NASA Annual Report, 1995

Project Title:
Next Generation Methods for Image Data Compression

GT Project Account Number E21-H86

Mark J. T. Smith, Professor
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

NASA Annual Report, December 12, 1995
Project Title: Next Generation Methods for Image Data Compression
GT Project Account Number E21-H86

Mark J. T. Smith, Professor
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

During this last year, we have continued our work in data compression and techniques that support image coding. The following papers have appeared in print during this time frame; preprints of those available were included in the last report.

1. F. Kossentini, M. Smith, and C. Barnes, "Image Coding Using Entropy-Constrained Residual Vector Quantization," *IEEE Transactions on Image Processing,* October 1995, pp. 1349-1358.

2. F. Kossentini, M. Smith, and C. Barnes, "Necessary Conditions for the Optimality of Variable Rate Residual Vector Quantizers," In *IEEE Transactions on Information Theory,* November 1995

3. F. Kossentini, W. Chung, and M. Smith, "Subband Image Coding with Jointly Optimized Quantizers," Proceedings of the *Int. Conf. on Acoustics, Speech, and Signal Processing*, Detroit, MI, 1995, pp. 2221-2224

4. W. Chung, F. Kossentini, and M. Smith, "A New Approach to Scalable Video Coding," *Proceedings of the Data Compression Conference*, March 1995.

5. Alen Docef, F. Kossentini, W. Chung, and M. Smith, "Multiplication-Free Subband Coding of Color Images," *Proceedings of the Data Compression Conference*, March 1995.

6. F. Kossentini, W. Chung, and M. Smith, "Subband Coding of Color Images with Multiplierless Encoders and Decoders," *Proceedings of the International Symposium on Circuits and Systems*, May 1995.

7. F. Kossentini, M. Smith and A. Scales, "High Order Entropy-Constrained Residual VQ for Lossless Compression of Images," Invited Paper in the *Proceedings of the International Symposium on Circuits and Systems*, May 1995.

8. F. Kossentini, W. Chung, and M. Smith, "Progressive Image Transmission Using Entropy Constrained Subband Coding," Invited paper in the *Proceedings of the International Conference on Digital Signal Processing*, June 26-28, Limassol, Cyprus, 1995.

During the course of this year, we have examined several new areas, which we found to be interesting: lossless compression extensions, progressive transmission extensions, and video coding extensions. Our papers describing these efforts are appended:

1. F. Kossentini, M. Smith and A. Scales, "High Order Entropy-Constrained Residual VQ for Lossless Compression of Images," Invited Paper in the *Proceedings of the International Symposium on Circuits and Systems*, May 1995.

2. F. Kossentini, W. Chung, and M. Smith, "Progressive Image Transmission Using Entropy Constrained Subband Coding," *Proceedings of the International Conference on Digital Signal Processing*, June 26-28, Limassol, Cyprus, 1995.

3. F. Kossentini, M. Smith, A. Scales, Tucker, "Medical Image Compression Using A New Subband Compression Method," *1995 SPIE Medical Imaging Conference*, San Diego, CA, February 26 - March 2, 1995.

4. A. Docef and M. Smith, "A Robust Model-Based Coding Technique for Ultrasound Video," *1995 SPIE Medical Imaging Conference*, San Diego, CA, February 26 - March 2, 1995, pp. 203-213

5. W. Chung, F. Kossentini, and M. Smith, "A New Approach to Scalable Video Coding," *Proceedings of the Data Compression Conference*, March 1995.

Progressive Image Transmission Using Entropy Constrained Subband Coding *

Faouzi Kossentini, Wilson C. Chung and Mark J. T. Smith
Digital Signal Processing Laboratory
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250

Abstract

A recently introduced subband coder with jointly optimized multistage residual scalar quantizers and entropy coders is employed in a progressive transmission environment. Both the subband and multistage residual structures are exploited to produce a completely embedded bit stream, making the subband coder suitable for progressive transmission applications. Different design methods involving both causal and non-causal encoding/decoding procedures are presented and evaluated subject to practical constraints. The high flexibility provided by multistage residual quantization and the effectiveness of the entropy coding strategy result in an attractive balance among reproduction quality, rate, resolution of progressive refinement, and complexity.

1 Introduction

In progressive compression systems, the decoder uses the incoming bits to reconstruct increasingly better reproductions of the signal being decoded. The user is thus afforded the capability to view a rendition of the image immediately, with picture quality improving dynamically as decoding is being performed. This has obvious benefits for telebrowsing and archival applications, rate-scalable codecs, and robust transmission applications over noisy channels.

Progressive transmission is related to the theory of "successive refinement of information," which was addressed by Equitz and Cover in [1]. They show that a rate-distortion problem is successively refinable if and only if the individual solutions of the rate distortion problems can be written as a Markov chain. It is also shown that while successive refinement is possible for some sources and distortion measures, it is not always achievable. Obviously, a source that is successively refinable can be sent progressively. However, progressive transmission does not necessarily imply that the source being transmitted is successively refinable.

The question that remains to be answered is: Given a successively refinable source (such as a memoryless Gaussian with squared error distortion measure), can we design a coder where the quantized source is transmitted progressively with no additional loss? The answer is yes

*This work was supported in part by the National Science Foundation under contract MIP-9116113 and the National Aeronautics and Space Administration.

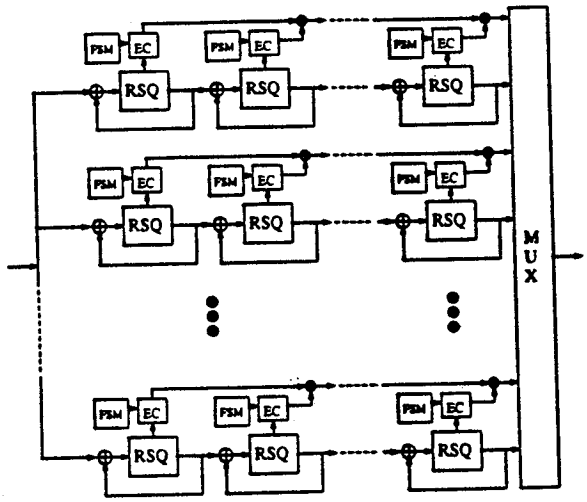


Figure 1: Basic block diagram of the subband encoder

if we had unlimited computing power and memory. However, given limited complexity, it is almost impossible to design sequential coders that achieve optimality. Thus, by imposing constraints supporting progressive image transmission, some additional distortion will generally be introduced to the final reconstructed image.

The subband coder upon which we build achieves excellent performance results [2], but also employs non-sequential encoding/decoding procedures. Much of this comes from the use of subband quantizers and entropy coders that are optimized jointly within and across stages and subbands. In this paper, we show how to include progressive transmission capability with only minimized loss in rate-distortion performance.

2 The Subband Coder

As shown in Figure 1, the input image is first decomposed into M subband signals using an analysis transformation. Each subband signal is then encoded using a sequence of P_m ($1 \leq m \leq M$) residual scalar quantization (RSQ) encoders. The output symbol of each of the stage quantizers is fed into an entropy coder driven by a high order stage statistical model that is governed by a finite state machine (FSM). The FSM allows the statistical model to switch between several first order (or zero order conditional) models by conditioning on the state of the FSM.

The algorithm used to design the subband coder minimizes iteratively the expected distortion, subject to a constraint on the complexity-constrained average entropy of the stage quantizers, by jointly optimizing the subband encoders, decoders, and entropy coders. The design algorithm employs the same Lagrangian parameter λ in the entropy-constrained optimization of all subband quantizers, and therefore requires no bit allocation [2].

The encoder optimization step of the design algorithm usually involves dynamic M-search of the multistage RSQ in each subband independently. The decoder optimization step consists of using the Gauss-Seidel algorithm [3] to minimize iteratively the average distortion between the input and the synthesized reproduction of all stage codebooks in all subbands. Since actual

entropy coders are not used explicitly in the design process, the entropy coder optimization step is equivalent to a potentially complex high order statistical modeling procedure. The multistage residual structure substantially reduces the large complexity demands, usually associated with FSM statistical modeling, and makes exploiting high order statistical dependencies much easier by producing multiresolution approximations of the input subband images. However, there are still many issues to be addressed. Multistage RSQs reduce the complexity because the output alphabet of the stage quantizers is typically very small (e.g., 2, 3, or 4), but the complexity of a stage entropy coder is still exponentially dependent on its order, or number of conditioning symbols or random variables, and/or the output alphabet sizes of the stage quantizers. Moreover, multistage RSQs also introduce another dimension to the statistical modeling problem, which significantly increases the number of possible combinations of conditioning symbols. Finally, many of the frequencies of combinations of conditioning symbols, gathered during the training process and used as estimates for probabilities, have zero values, producing empty states. This complicates the encoding stage because a combination of conditioning symbols corresponding to an empty state may occur. This is the so-called empty state problem, a problem usually associated with finite state machines. In [2], a complexity-constrained statistical modeling algorithm is proposed that attempts to simultaneously solve the above problems. Using overall complexity and average entropy of all stage quantizers as the criteria, the algorithm can be described as follows:

1. For each stage in each subband, and given a sufficiently large region of conditioning support, locate the best (in the sense of minimizing the entropy) 1st, 2nd, 3rd, ...order statistical models;

2. Substantially reduce the stage model orders, thereby reducing the number of conditioning states, by employing a tree structure and using the generalized BFOS algorithm;

3. Employ the well-known PNN algorithm to further reduce the number of conditioning states for each stage statistical model (state quantization).

3 Progressive Transmission

An important advantage of the subband coder introduced in [2] is its suitability for progressive transmission. The successive approximation nature of the subband structure results in multiresolution approximations of the input image. For example, a lowpass approximation can be obtained by transmitting information from only the low-low frequency band. Then, the quality of the reproduction can be successively improved by transmitting information from higher frequency bands. For this progressive transmission technique to be more effective, the subbands may have to be ordered in accordance with their perceptual importance. Whatever ordering is used, it must be the same in both design and encoding/decoding procedures. This insures that only previously coded symbols are required by the FSM. Fortunately, this type of progressive transmission places no constraints on the encoding/decoding performed within the subbands. Moreover, non-causal[1] encoding/decoding across subbands has negligible impact on performance anyway. However, incremental SNR improvements obtained between progressive updates tends

[1]Causality is defined in this paper in the context of operations across stages

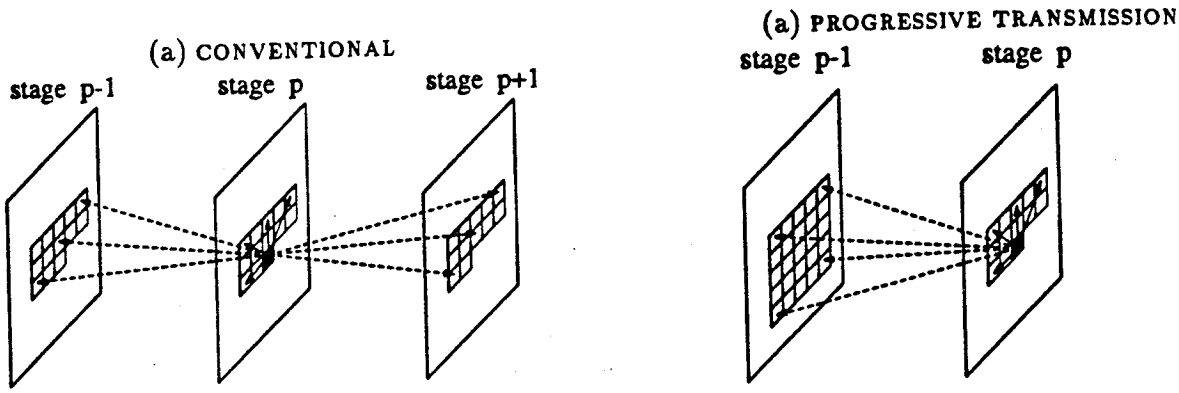


Figure 2: (a) Illustration of the conditioning structure used in previous work and (b) Illustration of the progressive transmission conditioning structure used in this work.

to be large. In fact, decoding the low-low band alone usually yields good quality. What we would prefer is to reconstruct a low quality initial image in a fraction of time, with increasing quality as bits are received.

To increase the resolution of the progressive transmission, we exploit the RSQ structure, which produces multistage approximations of the input image. That is, the farther down the quantization tree the input coefficient is encoded, the better, on average, the reproduction. Given that the image is decomposed into N subbands and RSQs with an average number of stages P are used to code the subband signals, the subband coder can enable the user to see a progressively improved image taken from a sequence of NP possible reconstructed images. In general, the subband/RSQ encodes the image using one stage codebook at a time until either transmission is aborted or until all NP stage codebooks have been used.

To use the RSQ for progressive transmission, the encoding and decoding processes have to be changed. First, the encoder must change the order in which bits are sent. Instead of sending all the bits for each input coefficient in each subband all at one time, the encoder scans a particular image subband and sends a small number of bits (which correspond to the stage codebook being used) for each coefficient. When encoding is repeated for a particular coefficient in a certain subband, the encoder recalls at which stage encoding stopped and continues encoding from there. Of course, the decoder changes its operation accordingly. Second, non-causal RSQ encoding/decoding can badly affect the performance in the intermediate stages. This is because we do not know when transmission is going to be halted. To insure that the quality of the intermediate reproductions is the best possible for the number of received bits, non-causal encoding such as dynamic M-search [4] and non-causal decoding such as applying the Gauss-Seidel algorithm to the stage codebooks have to be abandoned. Third, since the subband coefficients are coded and transmitted in a different order, conditioning must also be changed. Subsequent stage symbols of previously coded coefficients are no longer available, but non-causal spatial regions of support can now be used. Figure 2 shows a graphical illustration of a conventional conditioning scheme and another that supports progressive transmission. In figure 2(a), notice that half-plane support is present for conditioning at the current stage level p, and full-plane support is available for the previous stage level $p-1$. Having full-plane support in stages 1 to

$p-1$ allows the coder to exploit larger spatial dependencies, which are usually stronger than inter-stage dependencies.

4 Experimental Results

Several 512 x 512 USC database images were used for training. The image BOAT was not part of the training sequence and was kept for testing. Each image was decomposed into 16 uniform subbands. To initialize the design algorithm, a multistage RSQ is obtained for each subband image, as described in [2]. The number of scalars in each stage codebook is set to 3. A uniform stage codebook size was used for all RSQs in all subbands, which simplifies both quantization and arithmetic encoding/decoding. We determined empirically that 3-scalar stage codebooks provide the best tradeoff between resolution of progressive transmission, complexity, and performance for the training sequence.

Since full-resolution progressive transmission was being tested with conditioning structure as shown in Figure 2(b), dynamic M-search and joint decoding optimization were not performed. During the statistical modeling procedure, the total number of probabilities is set to a maximum of 512. For each state of the FSM model at each stage, only two probabilities, quantized to values between 0 and 255, are used by each adaptive arithmetic coder. Dynamic adaptation [5] was performed to further lower the bit rate.

The encoding/decoding complexity and memory of the fully embedded image subband coder are relatively small. The memory required to store all codebooks for each rate-distortion point is only 356 bytes, while that required to store the conditional probabilities is 512 bytes. Furthermore, the average number of multiplies/adds required for full-resolution encoding is approximately 10 per input sample. Decoding requires 6 multiplies/adds.

Figure 3 shows the test image BOAT coded at (a) 0.01 bits per pixel (bpp), (b) 0.07 bpp, (c) 0.16 bpp, and (d) 0.35 bpp. Notice that even that only the first stage indices in the low-low band were decoded in Figure 3(a), the image can still be recognized, and can be decoded rapidly.

References

[1] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. on Information Theory*, vol. 37, no. 2, pp. 269–274, 1991.

[2] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, (Detroit, MI, USA), Apr. 1995.

[3] F. Kossentini, M. Smith, and C. Barnes, "Necessary conditions for the optimality of variable rate residual vector quantizers," *Submitted to Transactions on Information Theory in June 1993. Revised in May 1994.*

[4] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114–116, July 1994.

[5] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Transactions on Communications*, vol. 29, no. 6, pp. 858–867, 1981.

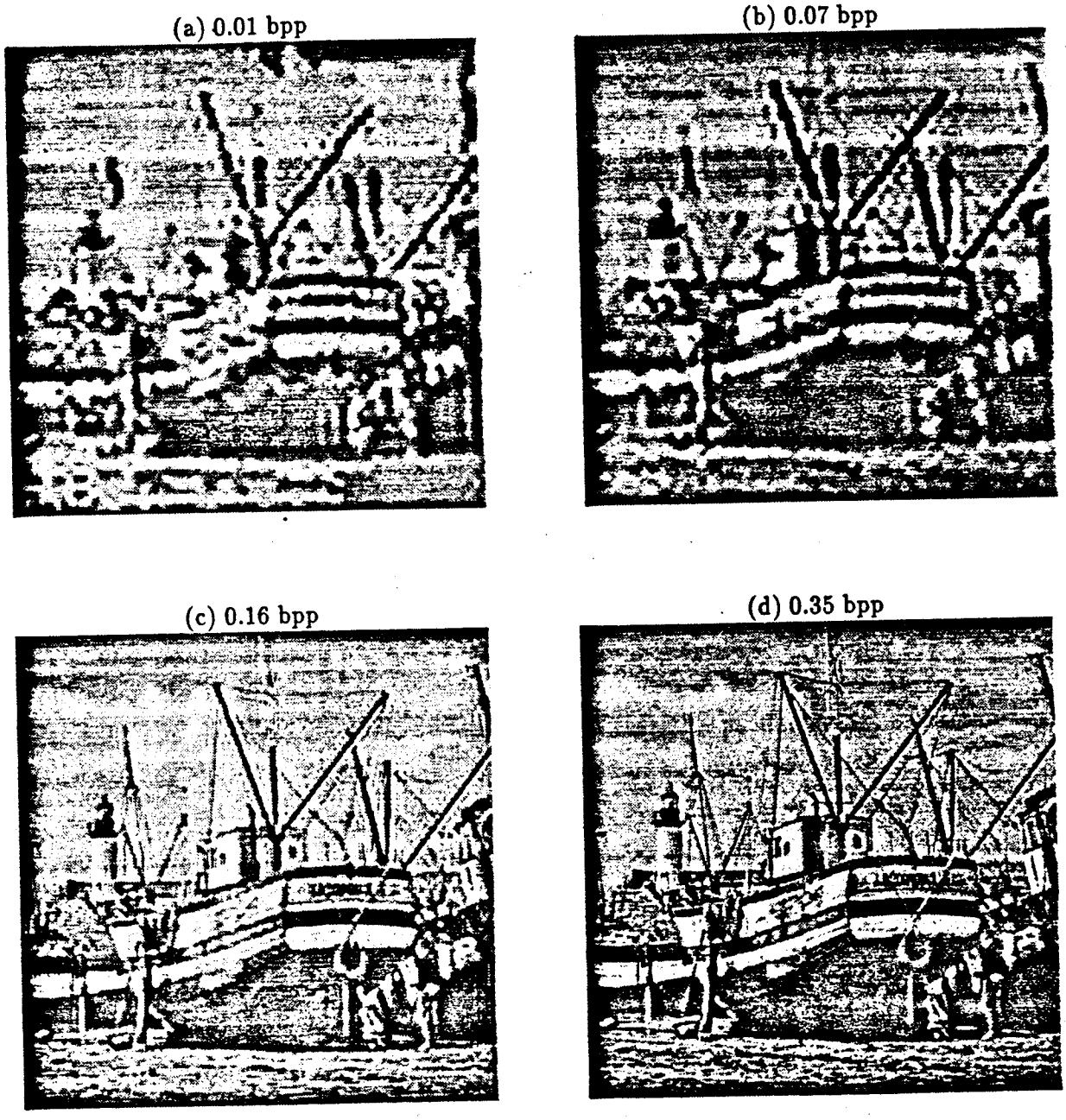


Figure 3: The image BOAT coded using the progressive transmission SBC at (a) 0.01 bpp, (b) 0.07 bpp, (c) 0.16 bpp, and (d) 0.35 bpp. The PSNRs (in dB) are 21.91, 23.30, 27.26, and 32.13, respectively.

HIGH ORDER ENTROPY-CONSTRAINED RESIDUAL VQ FOR LOSSLESS COMPRESSION OF IMAGES

Faouzi Kossentini and Mark J. T. Smith
School of ECE, Georgia Institute of Technology, Atlanta GA 30332-0250
Allen Scales
Nichols Research Corporation, MS 1-1-174, 4040 S. Memorial Pkwy, Huntsville AL 35802

Abstract—**High order entropy coding is a powerful technique for exploiting high order statistical dependencies. However, the exponentially high complexity associated with such a method often discourages its use. In this paper, an entropy-constrained residual vector quantization method is proposed for lossless compression of images. The method consists of first quantizing the input image using a high order entropy-constrained residual vector quantizer and then coding the residual image using a first order entropy coder. The distortion measure used in the entropy-constrained optimization is essentially the first order entropy of the residual image. Experimental results show very competitive performance.**

I. INTRODUCTION

A common approach to lossless image coding is to preprocess the data, in a way that removes statistical dependencies among the input symbols, and code those symbols with an entropy coder. Individual systems differ in their choice of statistical models for removing redundancies and their choice of entropy coders, like arithmetic and Huffman for example. Simple statistical models such as DPCM can remove some of the dependencies but usually are ineffective in handling high order dependencies.

High order statistical models have been proposed previously for lossless compression of binary images [1], and were shown to be very effective. Unfortunately, they cannot be translated efficiently to the gray-scale case. The computational and storage demands can be prohibitive. For example, a typical first order conditional statistical model might require that 65535 conditional probabilities be computed and stored. This number grows exponentially with increasing model order. Compounding the problem is the fact that many of the probability tables cannot be populated even when large training sequences are used, making high order entropy coding a very difficult task.

Several methods have been proposed recently for reducing the complexity of these statistical models [2, 3, 4, 5]. Most employ quantization or merging principles to reduce the number of conditioning states or tables of conditional probabilities, usually leading to orders of magnitude reductions in complexity while sacrificing only a small loss in performance. Others involve decomposition of the original signal into binary signals, which increases the accuracy of estimating the statistical model and thus improves the compression performance. In this paper, we introduce a new method that is based on both decomposition and probability table reduction techniques. Statistical modeling is performed through high order conditional entropy-constrained residual vector quantization (CEC-RVQ) [6, 7]. The entropy-based distortion measure employed in the CEC-RVQ optimization coupled with the high order entropy coding of the CEC-RVQ output result in substantial reductions in the entropy of the residual signal. This design framework, leads to high compression performance relative to other competing approaches.

II. PROPOSED FRAMEWORK

The hybrid technique of quantization and entropy coding of the residual signal has been shown to yield good compression performance [8, 9, 10]. This is due to the fact that quantization often produces a structure where high order statistical dependencies can be exploited. Moreover, since the output alphabet of the quantizer can be made smaller than that of the original signal, the complexity of high order statistical modeling is reduced. This is especially the case when structurally constrained quantizers are employed. In particular, the structure of the multistage residual vector quantization (RVQ) used here has been shown [11] to be very successful in providing more accurate estimates of the statistical dependencies of the original signal while also reducing drastically the complexity of high order statistical modeling. Multistage RVQ produces multiresolution approximations of the input signal, and allows high order statistical conditioning to be performed between the stage sub-signals.

As shown in Figure 1, we employ a CEC-RVQ to

This work was supported by the National Science Foundation under contract MIP-9116113 and the NASA.

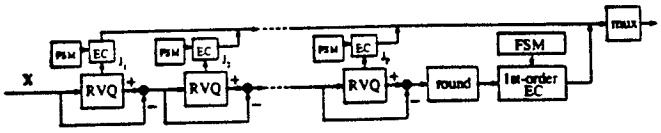


Figure 1: Proposed CEC-RVQ lossless coder.

quantize the input signal, where the output of the stage RVQ is then fed into a statistical-model-driven entropy coder (EC). The high order stage statistical model is represented by a finite-state machine (FSM) where the state transitions are based on previously coded symbols. The quantized signal is *rounded* to the nearest integer, and the residual signal, formed by subtracting the rounded quantized signal from the original one, is then coded using a first order entropy coder. Empirical work has shown that using higher order entropy coding does not lead to significant reductions in output entropy of the residual signal. In the final stage of the encoder, the bits emanating from the stage entropy coders as well as the residual entropy coder are combined together into a uniquely decodable bit stream, which is sent to the channel.

There are two important ideas, unique to this framework, that exemplify the novelty of this lossless approach. First, since the overall system is lossless, it is potentially better to employ the entropy of the residual signal as a distortion measure in the design of the CEC-RVQ. Using conventional distortion measures such as the squared error measure does not lead to minimization of the residual entropy. To elaborate, let $\boldsymbol{x}$ be the input and $\hat{\boldsymbol{x}}$ be the output of the CEC-RVQ. The new distortion measure used in the design of the CEC-RVQ is $d(\boldsymbol{x}, \hat{\boldsymbol{x}}) = -\log_2[\mathrm{pr}(I(\boldsymbol{x} - \hat{\boldsymbol{x}}))]$, where $I(\alpha)$ is the integer closest to the real α. The distortion is essentially the self-information of the integer-converted residual signal, and is used as an estimate of the length of the codeword that would be used to encode the symbol $I(\boldsymbol{x} - \hat{\boldsymbol{x}})$. In other words, the CEC-RVQ designed to minimize such a distortion measure also minimizes the entropy[1] of the residual signal.

The second idea is that only entropy is a measure of performance. Since the distortion measure is the entropy, the CEC-RVQ design algorithm produces an operational *entropy-entropy* curve where each point represents a pair of entropies, the first being the high order entropy h_o of the CEC-RVQ and the second being the entropy h_r of the residual. The high order entropy h_o is obtained by $h_o = \mathcal{H}(h_r)$, where $\mathcal{H}$ is the operational entropy-entropy function. It can be easily shown that the function $\mathcal{H}(h_r)$ is continuous and differentiable (except for some points). However, it is generally not convex, and its convexity depends on the source as well as the entropy measure used to estimate the information content in the residual signal. Fortunately, experimental work shows that for natural images and the first-order entropy, the function $\mathcal{H}(h_r)$ is convex with endpoints H_o and H, as illustrated in Figure (2). In the figure, the right endpoint H is the first-order entropy of the original signal. The left endpoint $H_o = \mathcal{H}(0)$ is the high order entropy of CEC-RVQ which results in perfect reconstruction after the CEC-RVQ output is rounded to the nearest integer. Due to the monotonicity of the CEC-RVQ (i.e., distortion will, on the average, only decrease by adding RVQ stages), H_o is finite. In other words, there is a point beyond which all of the real components of the residual signal lie in the real interval $(-0.5, 0.5)$. The problem at hand is to find an $(\mathcal{H}(h_r), h_r)$ pair such that the function $\mathcal{F}(h_r) = \mathcal{H}(h_r) + h_r$ is minimized. As shown in the figure, the minimum occurs at h_r^* such that $H'(h_r^*) = -1$. As will be shown later, the CEC-RVQ algorithm is based on a Lagrangian minimization where λ is the slope of the operational entropy-entropy function $\mathcal{H}$. Thus, the problem translates into designing the EC-RVQ with corresponding Lagrangian parameter lying in the neighborhood of 1.

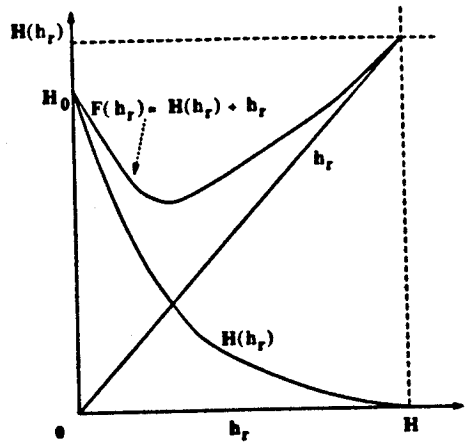


Figure 2: Illustration of an operational entropy-entropy curve.

Note that $\mathcal{F}$ would not necessarily have a minimum at h_r^* if $\mathcal{H}$ were not convex. Moreover, it is implied in Figure 2 that $H_o < H$. This is not true in general, since H depends on the source and H_o depends on the source, quantizer, and quantizer output statistical model. If $H_o \geq H$, the minimum may be larger or equal to the entropy H, and quantization becomes useless. However, by using CEC-RVQ, it is observed that H_o is usually significantly smaller than H. Thus, CEC-RVQ has the potential of achieving rates that are substantially lower than those obtained by first order entropy coding the original signal.

III. DESIGN AND COMPLEXITY ISSUES

The CEC-RVQ design algorithm proposed here iteratively minimizes the Lagrangian

$$J_\lambda = E[-\log_2 \mathrm{pr}(I(\boldsymbol{X} - \hat{\mathbf{X}}))] + \lambda E[\ell(\boldsymbol{L}(J|U))],$$

[1] This is the first-order entropy. For higher order entropies, high order probabilities should be used in the distortion measure.

where U is the state random variable [6], $\boldsymbol{L}$ is the high order conditional entropy mapping, and $\ell(\boldsymbol{L}(J|U))$ is the length of the variable length codeword $\boldsymbol{L}(J|U)$. The Lagrangian parameter λ controls the entropy-entropy tradeoffs and is used in the design process to locate on the operational entropy-entropy curve the point where the sum of the entropies is a minimum or close to a minimum.

In this work, a training sequence that is representative of the source output to be encoded is used in the design process. Let $\boldsymbol{x}^i$ be the ith k-dimensional vector taken from the training sequence of size N. An optimal encoding optimization step generally requires exhaustively searching the reproduction vector $\hat{\boldsymbol{x}}^*$ that minimizes the Lagrangian $-\log_2 \mathrm{pr}(I(\boldsymbol{x}^i - \hat{\boldsymbol{x}}^*)) + \lambda(-\log_2 \mathrm{pr}(j|u))$, where j is the current output of the CEC-RVQ and $u \in U$ is the current conditioning state. This typically yields large encoding complexity. To reduce complexity, non-exhaustive stage searching algorithms are usually used, leading to a good balance between complexity and encoding accuracy. In particular, the dynamic M-search algorithm [12], which is shown to generally perform better than the conventional M-search algorithm, is used here to search the CEC-RVQ.

The decoder optimization step consists of using the Gauss-Seidel algorithm [6] to iteratively minimize the average output entropy of the residual signal subject to fixed stage encoding partitions. Suppose the CEC-RVQ contains P stage VQ codebooks, each containing $N_p (1 \le p \le P)$ k-dimensional code vectors. Also, let $\mathcal{V}(j_p)$ denote the j_pth *non-causal* partition cell that corresponds to the j_pth code vector in the pth stage codebook. The partition cell $\mathcal{V}(j_p)$ is formed of all *stage-removed* residual vectors $\boldsymbol{\gamma}^i(j_p) = \boldsymbol{x}^i - \boldsymbol{z}_p^i(j_p)$, where $\boldsymbol{z}_p^i$ is given by

$$\boldsymbol{z}_p^i(j_p) = \sum_{l=1}^{p-1} \boldsymbol{y}_l(j_l^i) + \sum_{i=p+1}^{P} \boldsymbol{y}_i(j_i^i),$$

where $j_1^i, \ldots, j_P^i$ are the corresponding encoding decisions for the input vector $\boldsymbol{x}^i$. Each iteration of the Gauss-Seidel algorithm consists of sequentially replacing for each stage partition cell the old stage code vector $\boldsymbol{y}(j_p)$ with the centroid vector $\boldsymbol{c}(j_p)$ given by

$$\boldsymbol{c}(j_p) = \arg \min_{\boldsymbol{u} \in \mathbb{R}^k} \sum_{\boldsymbol{\gamma}^i(j_p) \in \mathcal{V}(j_p)} -\log_2 \mathrm{pr}(I(\boldsymbol{\gamma}^i(j_p) - \boldsymbol{u})). \quad (1)$$

The centroid vector $\boldsymbol{c}(j_p)$ is very difficult (if not impossible) to determine analytically. Thus, a numerical optimization procedure is used in this work. This further complicates the decoder optimization, but such iterative optimization is only performed in the design process and therefore does not affect the encoder/decoder complexity.

The entropy coder optimization consists of simply updating the finite-state machine (FSM) and the corresponding state tables of conditional probabilities [6]. Only the stage high order statistical models are optimized, and no actual entropy coders are embedded in the design loop. This simplifies the design process, but the complexity of the stage statistical models must still be addressed. Like VQ, high order statistical modeling provides a way to exploit high order statistics while also requiring complexity that is exponentially dependent on the parameters of the model. RVQ drastically reduces the complexity of the high order model and improves our estimates of the dependencies by generating multistage approximations of the input signal, where the output alphabets of the subspaces are small (e.g., 2, 3, or 4).

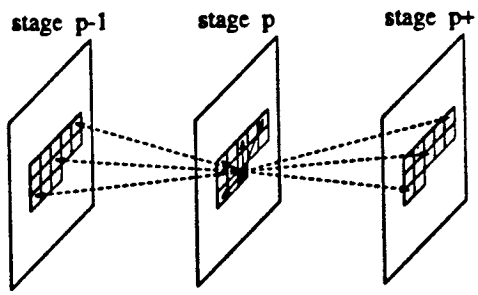


Figure 3: Illustration of a conditioning structure for CEC-RVQ.

Complexity-constrained statistical modeling for the output of the stage RVQs can be divided into three tasks. The first task is to locate a small number m_p of conditioning symbols (or previous outputs of some stage RVQs), given an initial region of support containing $\mathcal{R}_p$ conditioning symbols, such that the m_pth order conditional entropy is minimized. This is illustrated in Figure 3 for the case of image coding, where the shaded block in the middle is the stage vector upon which conditioning is being performed. A total of m (12 in this case) neighboring blocks is utilized for conditioning. These blocks define the spatial region supporting the conditioning. The solid arrows show these neighboring blocks at the pth stage. In addition to the spatial dimension, conditioning is based on corresponding blocks at different stage levels, which is illustrated in Figure 3 by the dashed arrows, showing these conditioning blocks at the $(p-1)$th and $(p+1)$th stages. By building a conditioning tree as described in [7] and using the dynamic M-search algorithm, one can find the best stage statistical models of orders 1, 2, 3, etc.

The second task to be performed is to determine the best orders for each of the stage statistical models subject to a constraint on overall complexity. For this purpose, a tree with P branches is built and populated with a sufficiently large number of complexity-entropy pairs in each branch. The well-known generalized BFOS algorithm [13] is then used to prune the tree to find the best stage orders subject to a limit T_1 on the number of conditional probabilities, used here as a measure for complexity.

Since relatively high orders are usually required to achieve a very low entropy, the complexity of the stage

IMAGE	HYBRID CODER	DPCM	[3]	[4]
LENA	4.27	4.80	4.42	4.20
BRIDGE	4.30	4.82	4.30	4.32

Table 1: Performance comparison of the hybrid lossless coder with DPCM, [3], and [4].

statistical models can still be high. Moreover, contextual information is usually located in a relatively small region of the state space. In other words, many states do not occur, and corresponding tables of conditional probabilities are not populated. Thus, the third task is to reduce the number of states while sacrificing a minimal loss in performance. The PNN algorithm [14] was shown to be successful in reducing the size of the stage statistical model by one order of magnitude while still limiting the increase in entropy to about 1%. The same approach used to locate the best stage statistical model orders is used here, where the PNN algorithm is applied to each of the stage statistical models with just-determined orders such that a new complexity-entropy pair is obtained every time two conditioning states are merged into a new one. The BFOS algorithm is again applied to identify the best numbers of conditioning states subject to a limit T_2 ($T_2 << T_1$) on the total number of conditional probabilities.

IV. EXPERIMENTAL RESULTS

Several images of size 512 × 512 taken from the USC database were used to design a CEC-RVQ codebook as described in the previous section. In all cases, test images were excluded from the training set. The CEC-RVQ codebook contains 12 stage codebooks with four 4×4 code vectors in each codebook. It is searched using the dynamic M-search algorithm, leading to approximately 60 vector Lagrangian calculations per input vector. The conditioning scheme we use is the one illustrated in Figure 3.

To locate the best orders for the stage models for a fixed maximum number of 4096 conditional probabilities, a balanced tree with depth 6 is constructed where the best $1, 2, \ldots, 6$ conditioning stage symbols are used. After the BFOS algorithm is employed, the number of conditioning states is further reduced by the PNN algorithm, whose outputs are used to populate yet another tree. Finally, the BFOS algorithm is used again to generate the FSM where the number of conditional probabilities is limited to 512.

The CEC-RVQ that yields the minimum overall entropy is determined as described previously using the training sequence. The corresponding set of stage codebooks, mapping tables generated by the PNN algorithm, and tables of conditional probabilities are used for encoding. Table 1 shows the entropy performance of the proposed hybrid coder, DPCM, and that of two of the best lossless compression techniques [3, 4] on the test images LENA and BOAT. The entropy is used as a measure so that the comparison is fair. An actual adaptive arithmetic coder was used to encode both the output of the stage RVQs and the residual image, and the compression ratios were slightly larger. Obviously, the proposed coder compares very favorably. Even better compression performance may be attained by using larger vector sizes and/or exploiting any statistical dependencies between the multistage images and the residual one. Preliminary experimental results are encouraging further study.

V. ACKNOWLEDGEMENT

The first author wishes to acknowledge W. Chung for valuable discussions.

REFERENCES

[1] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Trans. Comm.*, vol. 29, no. 6, pp. 858–867, 1981.

[2] S. M. Lei, T. C. Chen, and K. H. Tzou, "Subband HDTV coding using high-order conditional statistics," *IEEE JSAC*, vol. 11, pp. 65–76, Jan. 1993.

[3] X. Ginesta and S. P. Kim, "Semi-adaptive context-tree based lossless image compression," in *ICIP*, (Austin, TX), Nov. 1994.

[4] K. Popat and R. W. Picard, "Exaggerated consensus in lossless image compression," in *ICIP*, (Austin, TX), Nov. 1994.

[5] S. Yu, M. N. Wernick, and N. P. Galatsanos, "Lossless compression of multidimensional medical image data using binary-decomposed high-order entropy coding," in *ICIP*, (Austin, TX), Nov. 1994.

[6] F. Kossentini, W. Chung, and M. Smith, "Image coding using high-order conditional entropy-constrained residual VQ," in *ICIP*, (Austin, TX), Nov. 1994.

[7] F. Kossentini, W. Chung, and M. Smith, "Conditional entropy-constrained residual VQ with application to image coding," *Submitted to Trans. on Image Processing*, July 1994.

[8] J. Dozier and J. Tilton, "Data compression for data archival, browse or quick-look," in *Space & Earth Sci. Data Comp. Workshop*, (Snowbird, UT), Mar. 1991.

[9] J. T. D. Dovik and M. Manohar, "Compression through decomposition into browse and residual images," in *Space & Earth Sci. Data Comp. Workshop*, (Snowbird, UT), Mar. 1993.

[10] W. Abbott III, R. Kay, and R. Pieper, "Performance considerations for the application of the lossless browse and residual model," in *Space & Earth Sci. Data Comp. Workshop*, (Salt Lake City, UT), Mar. 1994.

[11] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," in *ICASSP*, (Detroit, MI), Apr. 1995.

[12] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114–116, July 1994.

[13] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on Information Theory*, vol. 37, pp. 400–402, Mar. 1991.

[14] W. H. Equitz, "New vector quantization clustering algorithm," *IEEE Trans. on ASSP*, vol. 37, pp. 1568–1575, Oct. 1989.

A SUBBAND CODING METHOD FOR HDTV

Wilson Chung, Faouzi Kossentini, and Mark J. T. Smith
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332

This paper introduces a new HDTV coder based on motion compensation, subband coding, and high order conditional entropy coding. The proposed coder exploits the temporal and spatial statistical dependencies inherent in the HDTV signal by using intra- and inter-subband conditioning for coding both the motion coordinates and the residual signal. The new framework provides an easy way to control the system complexity and performance, and inherently supports multiresolution transmission. Experimental results show that the coder outperforms MPEG-2, while still maintaining relatively low complexity.

I. INTRODUCTION

Several methods have been proposed recently for transmission of HDTV [1, 2, 3, 4, 5, 6, 7]. Most employ motion compensation at one stage or another, after which the residual between the original and predicted frames is computed and encoded spatially. DCT-based spatial coders are widely used, most notably in the MPEG standards. However, subband coders are also becoming popular.

There are many important issues that are associated with HDTV coding, such as control over the bit rate and picture quality, error correction and concealment, and multiresolution capability for multisource decoding and progressive transmission applications. In this paper, we introduce a new subband video coder which achieves good performance with low relative complexity, but also provides a framework where most of these issues can be easily addressed. The proposed coder employs motion estimation and compensation independently for each subband, but encodes the motion vectors using a high order conditional entropy coding scheme that exploits statistical dependencies between motion vectors of the same frame and successive frames as well as between the coordinates of the motion vectors, simultaneously. The coder also identifies non-compensatable blocks through the use of statistically optimized thresholding, which are then intra-frame coded. The video coder is described next. This is followed by a discussion of practical design issues. Section 4 presents experimental results which compare the performance and complexity of the coder with that of MPEG-2.

This work was supported by the National Science Foundation under contract MIP-9116113 and the National Aeronautics and Space Administration.

II. THE VIDEO CODER

First, consider a conventional subband video coder. In the parlance of MPEG, the frames that are coded spatially are called I frames. Those that are forward-predicted are called P frames, and those that are forward- and backward-predicted are called B frames. The sequence of video frames is first grouped into blocks of N frames, where the first frame (or I frame) is coded using an intra-frame subband coder, and the other $N-1$ frames (or P frames) are predicted using motion estimation and compensation, and the residual frames are coded using another subband coder. In this work, no B frames are used. At the receiver, each video frame is constructed from motion information (if applicable) and the coded residual frame.

There are two important problems associated with the above coder. First, motion compensation using the block matching algorithm with a typical block size of 16×16 and search range of -16 -to- $+16$ in each dimension is usually computationally intensive. This problem becomes even worse in HDTV coding because both block sizes and search areas have to be somewhat larger to achieve good performance. Second, due to the block matching algorithm, blockiness frequently appears in the residual frame, which introduces *artificial* high frequencies. To solve these two problems, we apply the block matching algorithm to each of the subbands. Figure 1 shows a block diagram of the proposed subband coder and Figure 2 shows the structure of the RVQ coder. Each frame is first decomposed into subbands using a tree-structured IIR analysis filter bank. The filter bank is based on two-band decompositions, which employ allpass polyphase separable IIR filters [8]. A full-search block matching algorithm (BMA) using the mean absolute distance (MAD) is used to estimate the motion vectors. Since the BMA does not necessarily produce the true motion vectors, we employ a thresholding technique for improving the rate-distortion performance. Let d_{min} be the minimum MAD associated with a block to be coded. Also, let T be a threshold, which is a large positive number empirically determined from the statis-

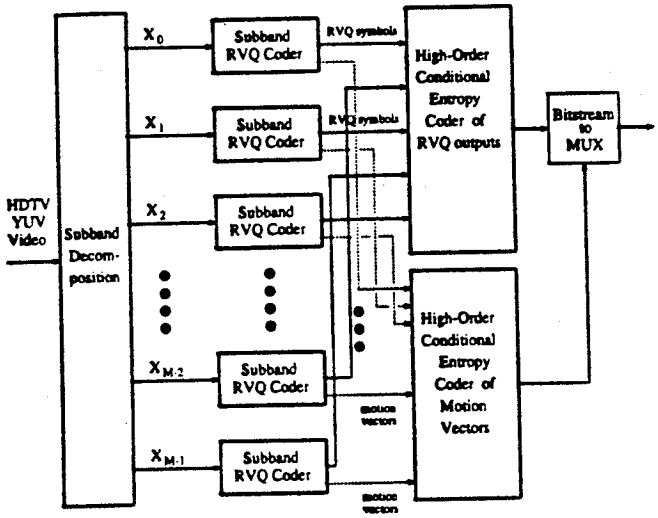


Figure 1: Proposed subband video coder.

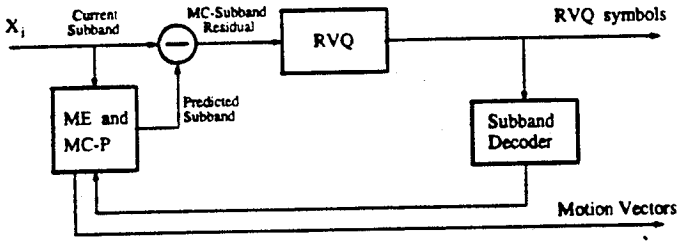


Figure 2: Basic structure of the RVQ coder.

tics of the subband being coded. If $d_{min} > T$, then the block is likely not compensatable. Thus, both the original block and the residual block, obtained by subtracting the motion compensated predicted block from the original one, are coded using the intra-band and residual coders, respectively, and the one leading to better rate-distortion performance is chosen (as will be described shortly). A special symbol, which can be coded as part of the motion information, is sent to the decoder indicating the type of coding used.

In many conventional HDTV subband coders as well as in MPEG, differential entropy coding of motion vectors is employed. Since motion vectors are usually slowly varying, the motion bit rate can be further reduced by exploiting dependencies not only between previous motion vectors within and across the subbands but also between the vector coordinates. For this purpose, we employ a high order conditional entropy coder that is based on finite state machine (FSM) modeling. More specifically, let $(X_{n,m}, Y_{n,m})$ be the pair of random variables representing the current horizontal and vertical motion displacements in the current subband (n,m) in frame n. Also, let $(U_{n,m}, V_{n,m})$ be the pair of state random variables with realizations $u_{n,m} = \{0,1,\ldots,S^u_{n,m}\}$ and $v_{n,m} = \{0,1,\ldots,S^v_{n,m}\}$, which we associate with $X_{n,m}$ and $Y_{n,m}$, respectively. Each state $u_{n,m}$ is given by

$$u_{n,m} = F_{n,m}(s^0_{n,m}, s^1_{n,m}, \ldots, s^{M_{n,m}}_{n,m}),$$

and each state $v_{n,m}$ is given by

$$v_{n,m} = G_{n,m}(t^0_{n,m}, t^1_{n,m}, \ldots, t^{N_{n,m}}_{n,m}),$$

where $s^0_{n,m}, s^1_{n,m}, \ldots, s^{M_{n,m}}_{n,m}$ and $t^0_{n,m}, t^1_{n,m}, \ldots, t^{N_{n,m}}_{n,m}$ are previously coded conditioning symbols. The mappings $F_{n,m}$ and $G_{n,m}$ are generally many-to-one mappings that convert combinations of realizations of the conditioning symbols to a particular state. Assuming that $X_{n,m}$ is entropy coded first, the conditioning symbols for the FSM model associated with $X_{n,m}$ are selected from a region composed of symbols located in all previously coded subbands (i.e., where motion vectors were already coded) in both frames n and $n-1$. When $Y_{n,m}$ is being coded, the horizontal displacements in the same subband can also be included in the conditioning region.

Statistical modeling for entropy coding the motion vectors consists of first selecting, for each subband (n,m), $M_{n,m}$ $(N_{n,m})$ conditioning symbols for $X_{n,m}$ $(Y_{n,m})$, and then finding mappings $F_{n,m}$ and $G_{n,m}$ such that the conditional entropies $H(X_{n,m}|U_{n,m})$ and $H(X_{n,m}|U_{n,m})$ are minimized subject to a limit on complexity. The total number of probabilities that must be computed and stored is used here as a measure of complexity. The tree-based algorithms described in [9] are used to find the best values of $M_{n,m}$ and $N_{n,m}$ subject to a limit C_1 on the total number of probabilities. The PNN algorithm [10], in conjunction with the generalized BFOS algorithm [11], is then used to construct mapping tables that represent $F_{n,m}$ and $G_{n,m}$ subject to another limit C_2 ($C_2 << C_1$) on the number of probabilities.

The intra-band (I-subband) and residual (P-subband) coders are multistage residual vectors quantizers (RVQs) followed with high order conditional statistical models, which are optimized to the intra-band and residual band statistics, respectively. Multistage RVQs provide an easy way to control the complexity-performance tradeoffs, and allow efficient high order statistical modeling. We restrict the number of code vectors per stage to be 2, which simplifies both statistical modeling and entropy coding used in this work. This also provides the highest resolution in a progressive transmission environment.

The same statistical modeling algorithm used for entropy coding the motion vectors is also used for entropy coding of the output of the RVQs. Both the motion vectors and the output of the RVQs are eventually coded using adaptive binary arithmetic coders (BACs) [12, 13]. These coders are very easy to adapt and require small complexity.

III. PRACTICAL DESIGN ISSUES

To achieve the lowest bit rate, the statistical models used to entropy code the motion vectors should be generated on-line. However, this requires a two-pass process where statistics are generated in the first pass, and the statistical modeling algorithm described above is used to generate the conditional probabilities. These probabilities must then be sent to the BAC decoders so that they can track the corresponding encoders. In most cases, this requires a large complexity. Moreover, even by restricting the number of states to be relatively small (such as 8), the side information can be excessive. Therefore, we choose to initialize the encoder with a generic statistical model, which we generate using a training HDTV sequence, and then employ dynamic adaptation [12] to track the local statistics of the motion flow.

For both the I-subbands and P-subbands, the multistage RVQs and associated statistical models are designed jointly using an entropy and complexity-constrained algorithm, which is described in [9, 14]. The design algorithm iteratively minimizes the expected distortion $E\{d(\boldsymbol{X}, \hat{\boldsymbol{X}})\}$ subject to a constraint on the overall entropy of the statistical models. The algorithm is based on a Lagrangian minimization and employs a Lagrangian parameter λ to control the rate-distortion tradeoffs. To substantially reduce the complexity of the design algorithm, only separate subband encoders and decoders are used. However, the RVQ stage encoders in each subband are jointly optimized through dynamic M-search, the decoders are jointly optimized using the Gauss-Seidel algorithm.

The most important part of the design algorithm is the encoding procedure, where either an intra-frame or inter-frame subband coder must be chosen for a particular block. Suppose we want to encode a block $B^i_{n,m}$ of size $L_{n,m}$ using the proposed I-subband and P-subband coders with Lagrangian parameters (or quality factors) λ_I and λ_P, respectively. The BMA algorithm is first applied, and the minimum MAD d_{min} is computed. If $d_{min} < T$, then the corresponding motion vector is encoded using the BAC specified by the current state, and the residual block is quantized using the P-subband (residual) RVQ. The output of each RVQ stage is encoded with a separate entropy coder composed of a FSM statistical model and a set of BACs, each specified uniquely by a state. If $d_{min} \geq T$, then the block is both I-subband and P-subband coded. Let $R_x = -\log_2 p(x^i|u^i)$ and $R_y = -\log_2 p(y^i|v^i)$ be estimates of the number of bits required to code the horizontal and vertical coordinates of the motion vector, respectively. Also, let d_P be the distortion and R_P be the rate that compose the minimum Lagrangian $J_P = d_P + \lambda_P R_P$ associated with coding the residual block. Assuming that $J_I = d_I + \lambda_I R_I$ is the minimum Lagrangian associated with coding the original block, then the I-subband coding method is selected if

$$J_I < d_P + \lambda_P(R_x + R_y + R_P).$$

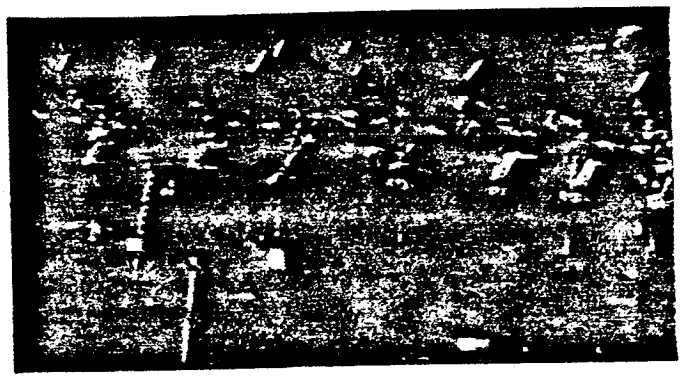


Figure 3: The 114th frame of the sequence BRITS

The proposed coder has many practical advantages, due to both the subband structure and the multistage structure of RVQ. For example, multiresolution transmission can be easily implemented in such a framework. Another example is error correction, where the more probable of the two stage code vectors is selected if an uncorrectable error is detected. Since each stage code vector represents only a small part of the coded vector, this should not significantly affect the reconstruction or the FSM statistical models.

IV. EXPERIMENTAL RESULTS

The image shown in Figure 3 is frame number 114 of the test sequence BRITS, which we encode using both the proposed coder and MPEG-2. The frame size is 720×1280. The original RGB color sequence with 8 bits/pixel requires approx 1.3 Gbs. The MPEG-2 software we used resides on *ftp.netcom.com:/pub/cfogg/mpeg2* [15].

In our experiments, each frame is decomposed into 64 uniform subbands, but more than half of the subbands are not coded. This is determined based on initial rate-distortion tradeoffs [9]. The BMA algorithm used in our experiments employs a block size of 2 × 2 and a search area of −2 -to- +2 in each dimension. Motion estimation is performed, and is done only for the Y luminance component and the estimated motion vector field is subsequently used for the motion compensation of U and V chrominance signals. A high order conditional entropy coder is designed for the motion vector coordinates, and one I-subband coder and one P-subband coder with vector size of 2 × 2 are designed for the each of the YUV components. We set the maximum allowed numbers of conditional probabilities for the motion entropy coder and the I-subband and P-subband entropy coders to $C_1 = 4094$ and $C_2 = 512$. The BACs used employ a *skew* factor between 1 and 256.

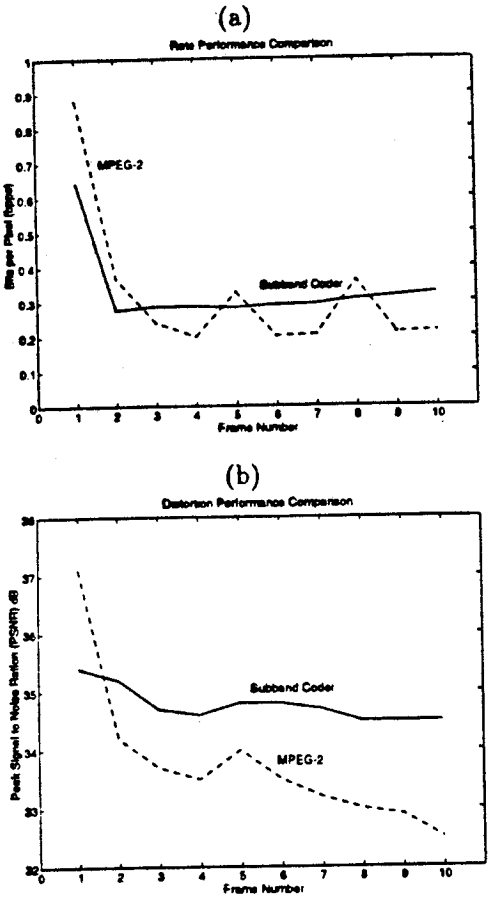


Figure 4: (a) Overall rate usage and (b) PSNR performance for the proposed coder.

For each rate-distortion point, the total memory required to store both the I-subband and P-subband RVQ codebooks and associated tables of conditional entropy codes is approximately 4.6 kilobytes. Moreover, only 512 bytes are required by the motion entropy coder. For analysis, quantization using dynamic M-search, and BAC encoding, approximately 27 multiplies and 32 adds per pixel are required. Only 3 multiplies and 14 adds are required for BAC decoding, inverse quantization, and synthesis. Not only are the encoding complexity and memory relatively small, but the performance is also good. Figure 4 (a) shows the average bit per pixel and Figure 4 (b) shows the PSNR result of our coder in comparison with the MPEG-2 standard for 10 frames of the luminance component of the color test video sequence BRITS. The average bit rate is approximately 18.0 Mbits/sec and the average PSNR is 34.75 dB for the proposed subband coder and 33.70 dB for MPEG-2. As is shown in the figure, the proposed coder clearly outperforms MPEG-2. Moreover, although MPEG-2 requires less encoding complexity and memory, the complexity of our subband coder are still reasonable.

REFERENCES

[1] J. G. Apostolopoulos and J. S. Lim, "Video compression for digital advanced television systems," in *Motion Analysis and Image Sequence Processing*, Boston: Kluwer Academic Publishers, 1992.

[2] S. M. Lei, T. C. Chen, and K. H. Tzou, "Subband HDTV coding using high-order conditional statistics," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 65–76, Jan. 1993.

[3] J. Bellisio and K. Tzou, "HDTV and the emerging broadband isdn network," *SPIE Proc. Visual Communications and Image Processing*, vol. 1001, pp. 772–786, 1988.

[4] F. Bosveld, R. Lagendijk, and J. Biemond, "Hierarchical coding of HDTV," *Signal Processing: Image Communication*, vol. 4, pp. 195–225, July 1992.

[5] F. Bosveld, R. Lagendijk, and J. Biemond, "Compatible HDTV transmission using conditional entropy coding," in *Proc. ICASSP*, vol. V, (Minneapolis, MN, USA), Apr. 1993.

[6] I. Furukawa, M. Nomura, N. Otha, and S. Ono, "Hierarchical coding of super high definition images with adaptive block-size multi-stage VQ," *Signal Processing of HDTV*, vol. III, 1992.

[7] S. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV," *IEEE Trans. on CVST* vol. 1, pp. 100–112, March 1991.

[8] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on ASSP.*, vol. 38, pp. 1446–1456, Aug. 1991.

[9] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on IP.*, July 1994.

[10] W. H. Equitz, "A new vector quantization clustering algorithm," *IEEE Trans. ASSP.*, pp. 1568–1575, October 1989.

[11] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on IT*, vol. 37, pp. 400–402, Mar. 1991.

[12] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Trans. on COM*, vol. 29, no. 6, pp. 858–867, 1981.

[13] W. B. Pennebaker, J. L. Mitchell, G. G. Langdon, and R. B. Arps, "An overview of the basic principles of the Q-coder adaptive binary arithmetic coder," *IBM J. Res. Dev.*, vol. 32, pp. 717–726, Nov. 1988.

[14] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," *Abstract Submitted to ICASSP*, (Detroit, MI, USA), Apr. 1995.

[15] MPEG Software Simulation Group, "MPEG-2 encoder / decoder, version 1.0," *ISO/IEC DIS 13818-2 codec*, May 1995.

Medical image compression using a new subband coding method

Faouzi Kossentini and Mark J. T. Smith

School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta GA 30332-0250

Allen Scales

Nichols Research Corporation
Huntsville AL 35815-1502

Doug Tucker

University of Alabama at Birmingham
Birmingham AL 35233

ABSTRACT

A recently introduced iterative complexity- and entropy-constrained subband quantization design algorithm is generalized and applied to medical image compression. In particular, the corresponding subband coder is used to encode computed tomography (CT) axial slice head images, where statistical dependencies between neighboring image subbands are exploited. Inter-slice conditioning is also employed for further improvements in compression performance. The subband coder features many advantages such as relatively low complexity and operation over a very wide range of bit rates. Experimental results demonstrate that the performance of the new subband coder is relatively good, both objectively and subjectively.

1 INTRODUCTION

Subband image coding of 8-bit/pixel natural images has been studied extensively in the literature.[1] Common to all subband image coding systems is the decomposition of the input image into subband images using a two-dimensional, mostly separable, filter bank. The resulting subband images are then quantized and entropy coded separately. Since the subband images typically have different statistical properties, a bit allocation algorithm is usually used to distribute bits among the subbands.

The subband image coder proposed in references[2,3] is different in that the design algorithm optimizes the subband quantizers and associated entropy coders jointly within and across the subbands in a complexity- and entropy-constrained framework. Advantages of the design algorithm are that it provides much greater control on the complexity-performance tradeoffs by using multistage residual vector quantizers,[4,5] and that no bit allocation

algorithm is required. This coder works very well for quasi-stationary signals such as most natural images. It can be designed to match the global statistics of a class of images by using a representative training sequence, and can be adapted to local statistics that are specific to individual images through the use of adaptive arithmetic coding.[6,7]

Like many natural images, medical images that are acquired from the same anatomical section using the same imaging modality are also quasi-stationary. A specific class of images, such as computed tomography (CT) axial slice head images, features similar global structural appearances due to the similarity in anatomical and tissue structures among different patients. On the other hand, anomalies such as pathologies or image artifacts, different density tissues, and different imaging conditions, produce image patterns that are not part of the training sequence. In medical imaging, some of these local statistics represent very critical information, and failing to reproduce such unique patterns can significantly impair the usefulness of the compressed medical image.

The problem of subband coding medical images using the coder proposed in references[2,3] is addressed in this paper. Although there are many similarities between natural and medical images, the problem of subband coding medical images is very different. Medical images are obtained from a variety of devices, and the images produced have different characteristics (e.g. dynamic range, spatial resolution) as well as distinct statistical dependencies. Performance can be improved by designing the filters, decomposition structures, quantizers, and entropy coders differently. For example, medical images contain a significant amount of both high and low frequency information. Thus, uniform decompositions fair better than the octave-band (or wavelet) decomposition frequently used in natural image subband coding. Moreover, a higher degree of fidelity is required in the compressed-decompressed images. Experimental work[8] shows that the choice of filters and filter design parameters has little or no effect on the reproduction quality in the low bit rate range. However, as will be discussed in this paper, filters do affect the subband coder's performance both objectively and subjectively in the high bit rate (high fidelity) range. Another problem associated with high fidelity subband coding is the large complexity usually required by the quantizers and corresponding entropy coders. Fortunately, since the proposed subband coder employs multistage residual vector quantizers, the complexity associated with both quantization and entropy coding is still relatively low.

The subband coder described in references[2,3] exploits both statistical intra-band and inter-band dependencies within an image simultaneously, mainly through complexity-constrained high order conditional entropy coding. In this work, inter-band dependencies both within a slice image and between slice images are exploited, resulting in a 5-10 % improvement in compression-complexity performance for the same reproduction quality. Next, we provide a brief description of the coder's components. This is followed by a discussion of design and complexity issues. This paper concludes with a discussion of the application of the subband coder to medical images and a presentation of some CT Head image coding experimental results.

2 THE SUBBAND CODER

Figure 1 shows the block diagram of the subband encoder used in this work. As is the case in conventional subband coding, the input image is first decomposed into M subband images using an analysis transformation. In this work, we employ a uniform tree-structured decomposition which is based on 2-band exact reconstruction filter banks. Each subband image is then encoded using a sequence of P_m $(1 \leq m \leq M)$ residual vector quantization (RVQ) fixed length encoders. Multistage RVQ is instrumental in drastically reducing the complexity of encoding/decoding as well as entropy coding, while still maintaining good rate-distortion performance. Advantages of multistage RVQs will be described in the following sections. Although encoding optimality can generally be achieved through exhaustive searching of the RVQ stage codebooks in all subbands (i.e. embedding the synthesis transformation in the encoding procedure), experiments have shown that dynamic M-search[9] of the stage codebooks in each subband separately usually leads to the best complexity/performance tradeoffs.

The output symbol of each of the stage vector quantizers is fed into an entropy coder driven by a high order

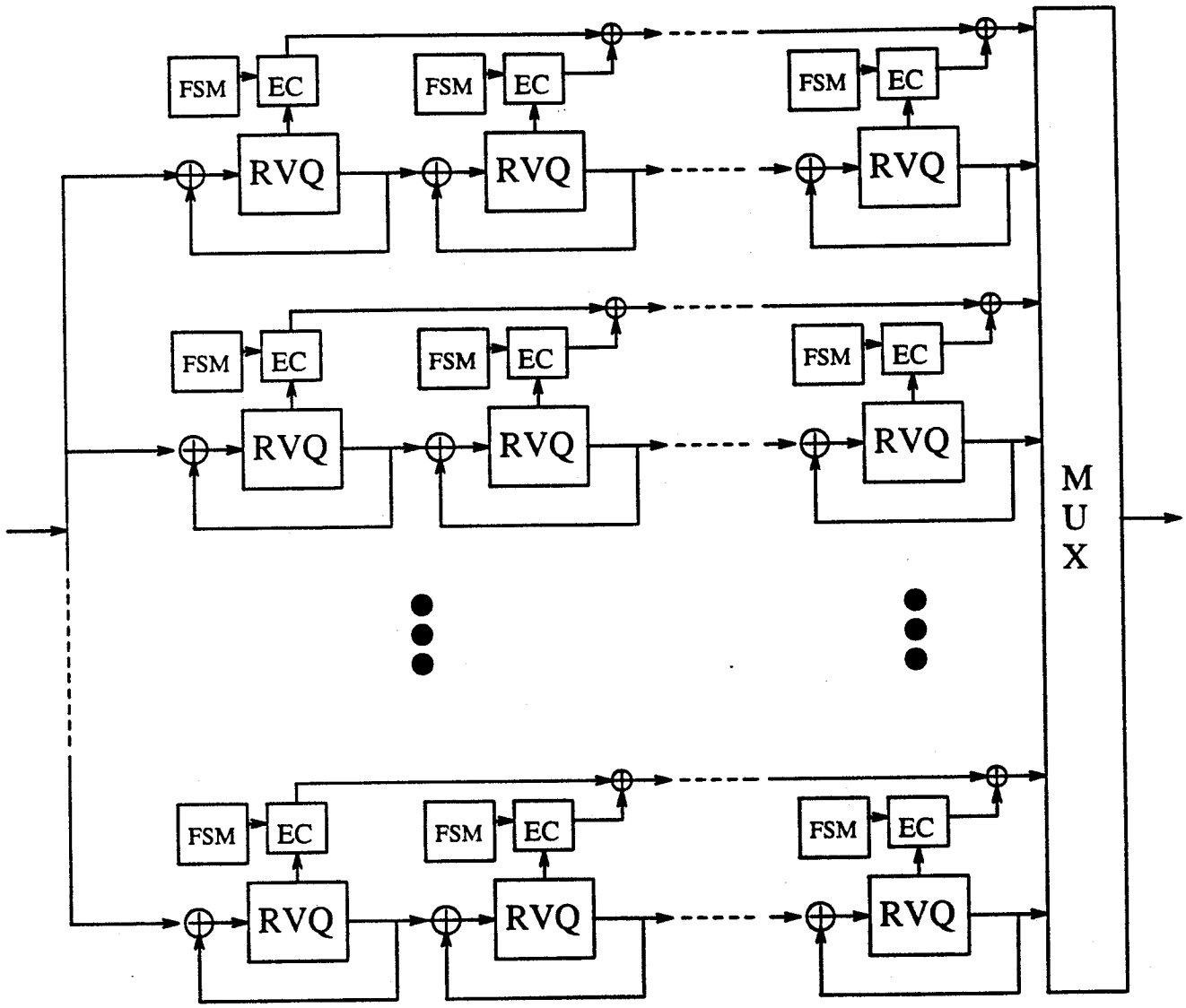


Figure 1: Basic block diagram of the subband encoder.

stage statistical model that is governed by a finite state machine (FSM). The FSM allows the statistical model to utilize information about previously coded stage vectors. A nonlinear function F given by $u = F(s_1, s_2, \ldots, s_n)$, where $s_1, s_2, \ldots, s_n$ are n conditioning symbols, or previous outputs of particular fixed-length RVQ stage encoders, is used here to determine the conditioning state u. As will be described in the next section, F is a many-to-one function that is represented by a table mapping each combination of realizations of conditioning random variables into a conditioning state. Since only previously coded symbols are used by the FSM, no side information is necessary and the decoder can track the state of the encoder by only storing the same table. Finally, the output bits of the entropy coders are combined together and sent to the channel.

3 DESIGN AND IMPLEMENTATION ISSUES

The algorithm used to design the subband coder minimizes iteratively the expected distortion subject to a constraint on the complexity-constrained high order conditional entropy of the stage vector quantizers (VQs). The popular squared error measure is used here as the distortion measure. This design algorithm is based on a Lagrangian minimization, and is a generalization of entropy-constrained algorithms described in.[10,11,5] Details of the *joint* optimality conditions used in the development of the algorithm and convergence issues are discussed elsewhere.[3]

Given a Lagrangian parameter, λ, which is chosen based on the overall rate and distortion of the subband system (i.e. no bit allocation algorithm is required), the entropy-constrained joint subband quantization algorithm consists of three optimization steps. The encoder optimization step involves exhaustively searching all RVQ stage codebooks, a task which requires a huge computational load. A large reduction in complexity can be achieved by using dynamic M-search. This results in only a small loss of performance. The decoder optimization step consists of using the Gauss-Seidel algorithm[5] to minimize iteratively the average distortion between the input and the synthesized reproduction of all stage codebooks in all subbands. The complexity can be drastically reduced by, for example, grouping neighboring stage codebooks in neighboring subbands and jointly optimizing each group independently. This typically results in less than a 0.10 dB loss in signal-to-noise (SNR) performance.

Since actual entropy coders are not used explicitly in the design process, the entropy coder optimization step is equivalent to a high order statistical modeling procedure. In terms of complexity (i.e. computational load and memory requirements), high order statistical modeling is potentially the most demanding task of the design algorithm. However, using the multistage residual structure not only substantially reduces the large complexity demands, usually associated with high order conditional entropy coding, but also makes exploiting high order statistical dependencies much easier by producing multiresolution approximations of the input subband images. However, there are still many issues to be addressed. Multistage RVQs reduce the complexity because the output alphabet of the stage quantizers is typically very small (e.g., 2, 3, or 4), but the complexity of a stage entropy coder is still exponentially dependent on its order (number of conditioning symbols or random variables) and/or the output alphabet sizes of the stage quantizers. Moreover, multistage RVQs also introduce another dimension to the statistical modeling problem, which significantly increases the number of possible combinations of conditioning symbols. Finally, many of the frequencies of combinations of conditioning symbols, gathered during the training process and used as estimates for probabilities, have zero values, producing empty states. This complicates the encoding stage because a combination of conditioning symbols corresponding to an empty state may occur. This is the so-called empty state problem, a problem usually associated with finite state machines.

In reference,[2] a complexity-constrained statistical modeling algorithm is proposed that attempts to simultaneously solve the above problems. To help illustrate the algorithm, Figure 2 shows the inter-stage, inter-band, and intra-band conditioning scheme employed in this work. Each image shown in the figure is a multistage approximation of a particular slice image. Note that statistical dependencies both within and across slice images can be exploited. For each stage (m,p) in each subband m, a 5-dimensional initial region of support $\mathcal{R}_{m,p}$ containing a sufficiently large number $R_{m,p}$ of conditioning symbols is first chosen. Then, the $n_{m,p}$, $n_{m,p} << R_{m,p}$, con-

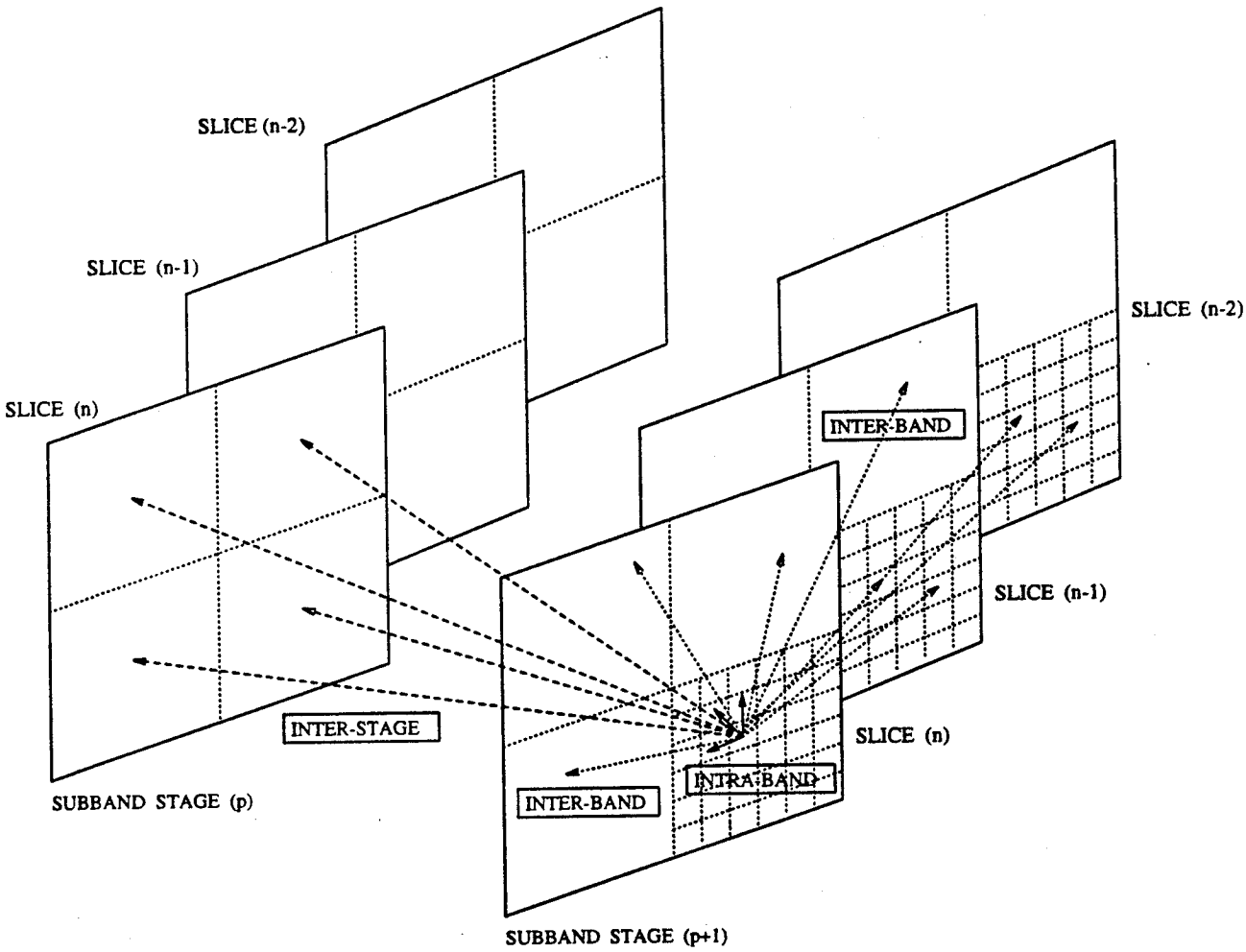


Figure 2: Inter-stage, inter-band, and intra-band conditioning scheme within an image sequence.

ditioning symbols $s^1, \ldots, s^{n_{m,p}}$ that lead to the smallest $n_{m,p}$th order conditional entropy $H(J_{m,p}|s^1, \ldots, s^{n_{m,p}})$ are located by building a special tree and using the dynamic M-search algorithm.

The next step of the algorithm is to find orders of all stage statistical models such that the average entropy in all subbands given a fixed level of complexity, expressed here in terms of total number of probabilities to be computed/stored, is minimized. The process described above is repeated for each stage (m,p) and many values of $n_{m,p}$, producing, let's say, $L_{m,p}$ complexity-entropy pairs per stage. For each complexity-entropy pair, complexity is given by $\mathcal{N}_{m,p} = S_{m,p}N_{m,p}$, where $S_{m,p}$ is the number of all combinations of realizations of the conditioning symbols and $N_{m,p}$ is the output alphabet size of stage p in band m. Once all complexity-entropy pairs are obtained, a tree with $\sum_{m=1}^{M} P_m$ branches, where P_m is the number of stage codebooks in the mth subband, can be built. Each branch of the tree is a unary tree of length $L_{m,p}$, and each node represents a complexity-entropy pair. The generalized BFOS algorithm[12] is then used to minimize the average entropy subject to a constraint $\mathcal{N}_1$ on the total number of conditional probabilities.

The FSM statistical model for each stage (m,p) employs a mapping F to determine the state given $n_{m,p}$ available conditioning symbols. This mapping F is one-to-one and is actually given by a table that contains the numbers $0, 1, \ldots, S_{m,p} - 1$, representing each of the possible combinations. Up to this point, the number of conditioning states for each of the stages is typically large. Many of the corresponding tables of probabilities may still be empty after the design is completed, thereby occupying memory which can usually be more efficiently used. Moreover, as mentioned earlier, these empty tables may be visited during actual encoding even though they were never visited during the design process. Therefore, the last step of the algorithm is to further reduce the number of conditioning states through quantization. In this work, the PNN algorithm[13] has been shown to be successful in reducing the number of states by orders of magnitude while still bounding the loss in entropy performance to about 1%. The PNN algorithm first merges all of the empty states with the least probable state into *one* conditioning state, thereby completely removing empty states. Then, the two conditioning states resulting in the lowest increase in entropy (when merged) are combined into one conditioning state, and so on until only one state, which represents one table of first order probabilities, is obtained. Since the objective is to minimize the complexity-constrained average entropy, the BFOS algorithm is again used, where a much smaller complexity value $\mathcal{N}_2$ is the constraint or criterion.

In the context of medical image compression, quantization of the conditioning states has two important advantages. First, the stage statistical model orders can be allowed to grow to relatively large numbers, which generally results in significantly lower average entropy because most medical images feature high order global statistical dependencies. This also incurs a small additional enoding/decoding complexity, since only larger mapping tables have to be stored/accessed. Second, the merging process improves the robustness of the subband coder because only global statistics are carried through, and the possibility of a strong mismatch between individual medical images and the subband coder is less likely.

4 EXPERIMENTAL RESULTS

A total of 90 axial slice CT head images with no abnormal findings were used for training and testing. Images were selected retrospectively from studies of 10 patients undergoing scans as a part of their clinical care. A General Electric (Waukasha, Wi) Hi-Lite Advantage CT scanner was used to produce all images which were either 3 mm (posterior fossa, 120 kVp, 320 mAs) or 5 mm (mid-brain, 120kVp, 240 mAs) thick slices. All images were of size 512 x 512 with 12 bit/pixel amplitude resolution. No special image processing or reconstruction algorithms were applied. Each image was extracted from the CT scanner's proprietary database using software tools supplied by the manufacturer. Subsequently, the proprietary header information was removed and the raw images were stored with 16-bit amplitude precision. A set of 12 slice images was kept for testing, and was not used as part of the training sequence.

Two experiments were performed. The first investigated the performance of the spatial subband coder, while the second considered exploiting both spatial and inter-slice dependencies within the CT image sequence. In both experiments, each slice image was first fed into a 2-level balanced tree structured filter bank, producing 16 image subbands. The A11 allpass polyphase exact reconstruction IIR filters[14] were used in our simulations. Many other filters, such as the Johnston 16-tap and 32-tap QMFs[15] and the Daubechies 32-tap wavelet filters,[16] were tested and were found to be inappropriate. The SNR reconstruction performance of these filters for the test CT images did not exceed 52 dB even when quantization was not performed. This is unsatisfactory in light of the fact that the medical community demands a SNR reconstruction performance that is usually 50 dB or higher.

In this work, we employ a vector size of 1×1 (scalar quantizer). Although k-dimensional vector quantizers are potentially better than scalar quantizers, their complexity is very large. Thus we found scalar quantizers to be more appropriate, particularly considering the high rates of operation. To initialize the design algorithm, a multistage residual scalar quantizer (RSQ) is obtained for each subband image, as described in reference.[3] The number of scalars in each RSQ stage codebook is set to 3. Non-uniform stage codebook sizes were considered, but no significant improvement in rate-distortion performance was obtained. Furthermore, choosing a uniform stage codebook size for all RSQs in all subbands simplifies both quantization and arithmetic encoding/decoding. We have also tried stage codebooks of sizes $2, 3, 4, 5, \ldots$ and have determined that 3-scalar stage codebooks provide the best complexity-performance tradeoffs for the training CT sequence.

In both experiments, dynamic M-search with a fixed threshold of 10 was used in the encoder optimization. Moreover, a joint decoder optimization between stages only is used in both cases. In other words, no joint optimization between subband decoders is performed. During the statistical modeling procedure, the value of $\mathcal{N}_1$ was set to 8192, and the value of $\mathcal{N}_2$ was set to 1024. For each state of the FSM model at stage (m, p), only two probabilities, quantized to values between 1 and 256, are needed by each adaptive arithmetic coder. Since the probabilities are constrained to be powers of 2, no multiplications are necessary in the implementation of the arithmetic encoders/decoders. Dynamic adaptation[17] was performed to further lower the bit rate. Although good performance high rate coders typically require a large design complexity, such is not the case in the first experiment. About 12 CPU hours on a Sparc 10 Sun Station were required to design subband coders operating at rates between 0.80 and 2.0 bpp. However, the design complexity in the second experiment is relatively large. More specifically, more than two days in CPU time were required to design the same number of codebooks and corresponding entropy coders. This is due to the fact that inter-slice conditioning requires that a much larger region of support be used, which complicates statistical modeling.

The encoding/decoding complexity and memory of the CT image subband coder are relatively small. The memory required to store all codebooks for each rate-distortion point is only 1152 bytes, while that required to store the conditional probabilities is approximately 1024 bytes. Furthermore, the average number of operations (multiplies/adds) required for encoding is 14.64 per input sample. Decoding requires 10 multiplies/adds. By placing some constraints on the coder, encoding/decoding can also be implemented without multiplications. However, such constraints also affect the rate-distortion performance. Full evaluation of a multiplication-free implementation of this subband coder for medical image compression is the subject of further research.

The objective quality of the reconstructed CT slice images is very good. Table 1 shows rates and SNRs for all 12 test CT slice images for average rates of 2.00, 1.50, 1.0 and 0.80 bpp, corresponding to compression ratios of $6:1$, $8:1$, $12:1$, and $15:1$, respectively. The SNR is defined by

$$\mathrm{SNR} = -10 \log_{10} \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (x(i,j) - \hat{x}(i,j))^2}{\sum_{i=1}^{N} \sum_{j=1}^{M} (x(i,j) - \mu)^2} \tag{1}$$

where $N \times M$ is the number of samples in the image, $x(i,j)$ and $\hat{x}(i,j)$ represent the original and the coded value (respectively) of the (i,j)th sample, and μ is the mean of $x(i,j)$. Figure 3(a) shows the original slice image # 11. Figure 3(b) shows the residual image formed by taking the absolute difference between the original image and the reconstructed one at a bit rate of 0.73 bpp. Note that the intensities of the residual image have been magnified by a factor of 16. Finally, Table 2 compares the bit rates and SNRs of the first and second experiments for the

	6:1		8:1		12:1		15:1	
	BR	SNR	BR	SNR	BR	SNR	BR	SNR
SLICE #1	2.16	56.83	1.75	52.89	1.13	48.75	0.89	46.35
SLICE #2	2.12	57.06	1.61	52.99	1.12	48.64	0.92	46.50
SLICE #3	2.09	57.11	1.67	53.05	1.16	48.77	0.89	46.41
SLICE #4	2.11	56.97	1.59	52.89	1.05	48.61	0.83	46.24
SLICE #5	1.99	57.27	1.42	53.00	0.98	48.90	0.79	46.62
SLICE #6	2.01	57.13	1.46	52.92	1.01	48.83	0.80	46.59
SLICE #7	2.03	57.18	1.48	52.94	1.02	48.82	0.82	46.54
SLICE #8	2.01	57.31	1.45	53.03	0.95	48.89	0.80	46.62
SLICE #9	1.94	57.35	1.39	53.10	1.02	48.86	0.77	46.61
SLICE #10	1.89	57.33	1.42	53.03	0.98	48.84	0.74	46.58
SLICE #11	1.84	57.36	1.38	53.06	0.93	48.89	0.73	46.65
SLICE #12	1.84	57.13	1.33	52.88	0.90	48.93	0.71	46.

Table 1: Bit rate (BR) in bits per pixel (bpp) and signal-to-noise ratio (SNR) in decibels (dB) for the 13 slice images used in the first experiment at compression ratios of 6:1, 8:1, 12:1, and 15:1.

	6:1		15:1	
	BR	SNR	BR	SNR
Non-inter-slice	1.84	57.36	0.73	46.65
Inter-slice	1.71	57.29	0.68	46.68

Table 2: Bit rate (BR) in bits per pixel (bpp) and signal-to-noise ratio (SNR) in decibels (dB) for the slice image #11 at compression ratios of 6:1 and 15:1.

slice image # 11 at the two 6 : 1 and 15 : 1 compression ratios. Looking at Table 2, one can see that inter-slice conditioning resulted in a 7 % decrease in bit rate roughly for approximately the same objective quality.

Compressed and reconstructed images of the non-inter-slice conditioning experiment were also viewed by an experienced radiologist for his impressions. Viewing was performed in a low-light environment. Images were displayed on an Image Systems M21PMAX 1280 × 1024 display using a Dome MD2kEISA display controller on a DELL Omniplex Pentium personal computer running MS-DOS 6.21 and an image viewing software customized from DOME's software library. The same series of 12 images, each compressed at 6:1, 8:1 and 12:1, were used. For each image viewing, the original image and a single compressed-reconstructed image were displayed together. All images compressed at 6:1 were viewed first, followed by the 8:1 and the 12:1 images, respectively. The radiologist was allowed to adjust window and level settings and no time constraints were imposed. The radiologist's impression was solicited. The radiologist reported no noticeable difference between the original image and the 6:1 or 8:1 compressed-reconstructed images. For two of the twelve 12:1 compressed-reconstructed images the observer noted slight enhancement of the high frequency component (noise) of the compressed-reconstructed image.

5 CONCLUSIONS

The results of the preliminary viewing of the compressed-reconstructed images by a radiologist were encouraging. We are currently conducting more rigorous observer performance tests to determine objectively the performance of radiologists using the compressed-reconstructed images. The computational complexity and memory requirements make this coder a suitable candidate for implementation in real-time hardware.

6 ACKNOWLEDGMENTS

The authors would like to acknowledge J. Kevin Smith, M.D., Ph.D. for his assistance. The first author would also like to acknowledge Wilson C. Chung for supplying software implementations for the various filters studied in this paper.

7 REFERENCES

[1] J. W. Woods, ed., *Subband Image Coding.* Norwell, MA: Kluwer Academic Publishers, 1991.

[2] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, (Detroit, MI, USA), Apr. 1995.

[3] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on Image Processing*, July 1994.

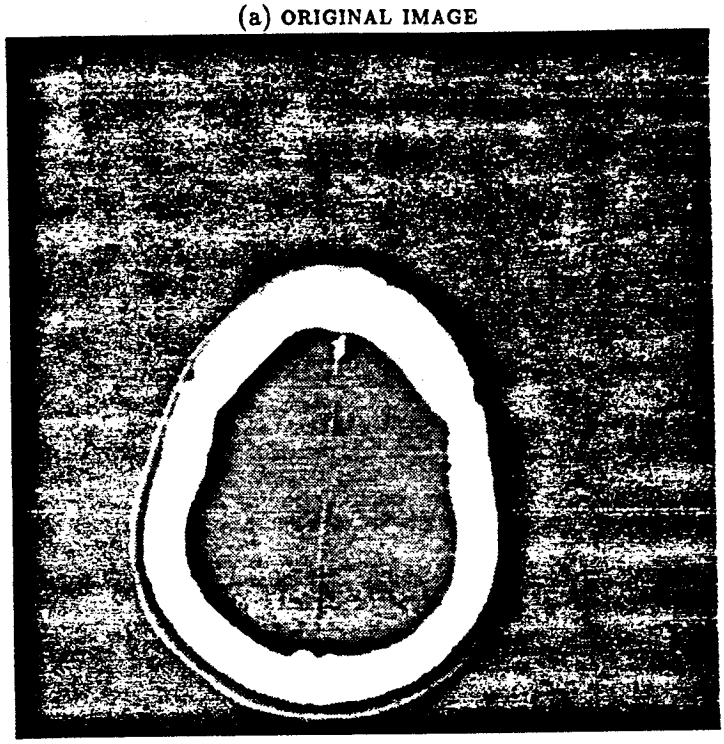
[4] B. H. Juang and A. H. Gray, "Multiple stage vector quantization for speech coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 597–600, April 1982.

[5] F. Kossentini, M. Smith, and C. Barnes, "Necessary conditions for the optimality of variable rate residual vector quantizers," *Submitted to Transactions on Information Theory in June 1993. Revised in May 1994.*

[6] G. Langdon, "An introduction to arithmetic coding," *IBM J. Res. Dev.*, vol. 28, pp. 135–149, Mar. 1984.

[7] A. C. Popat, *Scalar Quantization with Arithmetic Coding.* PhD thesis, M.I.T., Cambridge, MA, 1986.

(a) ORIGINAL IMAGE



(b) RESIDUAL IMAGE

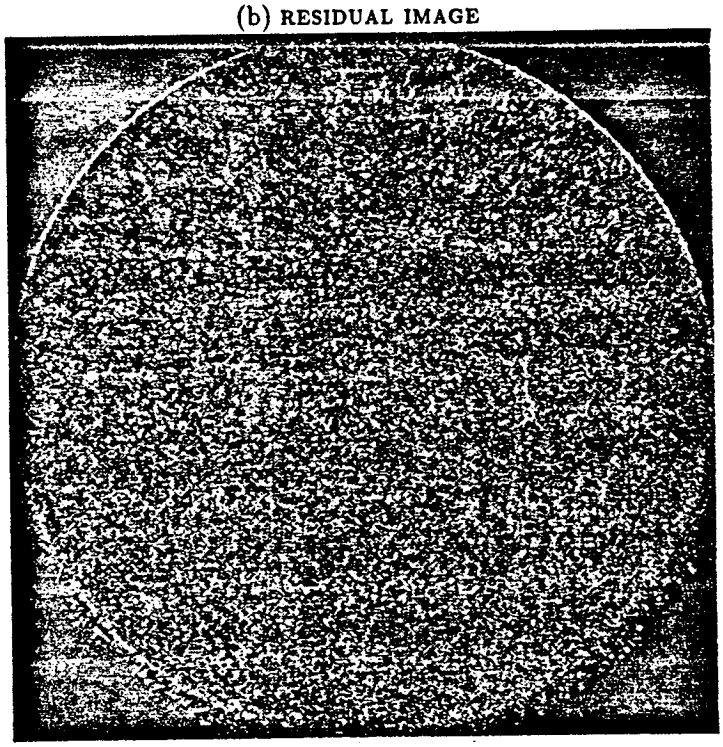


Figure 3: The original slice image #11 and the corresponding residual image at a bit rate of 0.73 bpp. Intensities of the residual image have been magnified by a factor of 16.

[8] A. Docef, F. Kossentini, W. Chung, and M. Smith, "Multiplication-free subband coding of color images," in *Data Compression Conference*, (Snowbird, UT, USA), Mar. 1995.

[9] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114–116, July 1994.

[10] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-37(1), pp. 31–42, Jan. 1989.

[11] F. Kossentini, M. Smith, and C. Barnes, "Entropy-constrained residual vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. V, (Minneapolis, MN, USA), pp. 598–601, Apr. 1993.

[12] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on Information Theory*, vol. 37, pp. 400–402, Mar. 1991.

[13] W. H. Equitz, "New vector quantization clustering algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1568–1575, Oct. 1989.

[14] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1446–1456, Aug. 1991.

[15] J. Johnston, "A filter family designed for use in quadrature mirror filter banks," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pp. 291–294, April 1980.

[16] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, Pennsylvania: SIAM, 1992.

[17] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Transactions on Communications*, vol. 29, no. 6, pp. 858–867, 1981.

A Robust Model-Based Coding Technique for Ultrasound Video *

Alen Docef and Mark J. T. Smith

Georgia Institute of Technology, School of Electrical and Computer Engineering
Atlanta, Georgia 30332-0250

ABSTRACT

This paper introduces a new approach to coding ultrasound video, the intended application being very low bit rate coding for transmission over low cost phone lines. The method exploits both the characteristic noise and the quasi- periodic nature of the signal. Data compression ratios between 250:1 and 1000:1 are shown to be possible, which is sufficient for transmission over ISDN and conventional phone lines. Preliminary results show this approach to be promising for remote ultrasound examinations.

Keywords: telemedicine, ultrasound, video coding, model-based coding, subband coding.

1 INTRODUCTION

Ultrasound video is a very cost effective diagnostic modality, and thus is widely used throughout this country and the world. Although ultrasound equipment is often available in rural and remote corners of the country, specialists to interpret data are typically in short supply in these locations. With the interest and support in telemedicine, the notion of having specialists perform ultrasound examinations at remote locations via electronic data exchange is very attractive. In the absence of channel bandwidth constraints, such an approach is straightforward, with high potential benefits related to providing immediate care and lowering overall expense. Unfortunately, many of these remote locations do not have access to or cannot afford to use high capacity channels (such as T1 lines) to interface with large well-staffed urban medical centers where such specialists reside.

In the presence of channel bandwidth constraints, this approach is encumbered by the large volume of data associated with digital video. Effective compression of the ultrasound prior to transmission will allow this data transfer to occur. The key is to achieve sufficient compression with acceptable reconstruction quality at rates compatible with telephone and ISDN lines. In this work we consider ultrasound video of the heart, where the remote examination involves a specialist at a remote location guiding the attending practitioner by telephone. A critical part of this examination is obtaining proper positioning of the ultrasound probe, so that a diagnosis can be made. The transmitted video quality standards for positioning purposes are clearly not as stringent as those for diagnosis. If positioning quality can be achieved, then higher quality video can be transmitted in a non-real time mode for diagnosis. Of course we hope to eventually be able to transmit diagnostic quality ultrasound in real time, but this not yet in reach. Regardless, the approach outlined above is a marked improvement in terms of accuracy and speed over sending video tapes by courier.

*This work was supported in part by the National Science Foundation under contract MIP-9116113 and by NASA.

The target goals imply compression ratios in the range from 250:1 to 1000:1. An obvious first line of attack on this problem is to investigate to what extent spatial and temporal sampling (i.e. frame size and frame rate) can be decimated without significantly impairing the quality. This has the advantage of being attractive computationally. Based on feedback from the Medical College of Georgia, a 4:1 reduction in spatial resolution to a size of 256×256 was judged to be acceptable. However, the full 30 f/s frame rate was recommended, particularly for pediatric cardiology where the heart rates are often very high.

Conventional coding methods such as H.261 and MPEG are not well suited to ultrasound video. The data rates tend to be too high and they have difficulty representing the high frequency information in the input. Model-based methods on the other hand are known for high compression ratios but suffer typically from variegated performance behavior over a wide variety of inputs.

In this paper, we introduce a model-based method that provides both high compression and robust behavior. To meet the difficult compression requirements imposed by the telephone bandwidth, it is important to identify and exploit all available properties of the signal and preserve with fidelity those parts of the signal that are important for expert analysis. In the case of ultrasound video in cardiology, for example, cardiologists must be able to see the shape of the walls, the shape and thickness of the valves and the tissue texture. By taking into account the nature of the noise/texture associated with the ultrasound images and identifying the important components (wall boundaries, valves, etc.), we formulated a visual model that can be used for very low bit rate coding.

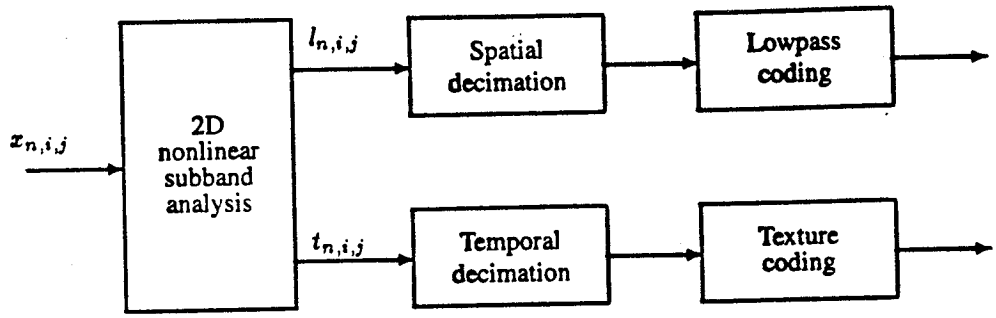
2 SYSTEM DESCRIPTION

The components of the proposed coding system are outlined in Figure 1. First, each input frame $x_{n,i,j}$ is decomposed nonlinearly into two components: a lowpass component, which is denoted by $l_{n,i,j}$; and a highpass or texture component, $t_{n,i,j}$, where n is the frame number, i is the row number, and j is the column number. The decomposition is based on a signal model and is optimized empirically such that the lowpass component contains most of the information needed for diagnosis, such as the contours of walls and valves. The highpass component contains information about the texture of the tissue being examined. The non-linear subband decomposition (upon which we elaborate later) is shown as the first block in Figure 1. After the decomposition, the lowpass component is then decimated in i and j to the Nyquist rate. Signal coding is then performed using an optimized subband coding method recently developed in the digital signal processing laboratory at Georgia Tech.[1] Some details of this method are presented in a later section.

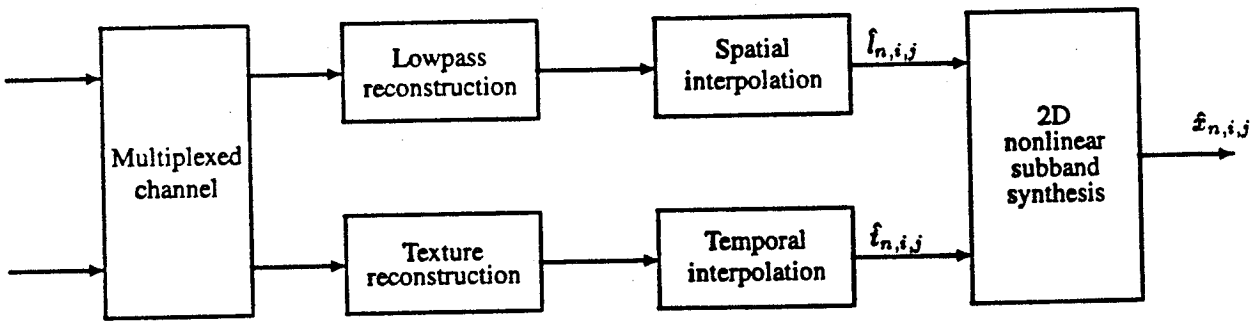
The lower branch of the system contains the texture information. It is decimated in the *temporal* domain and encoded using an in-house version of entropy-constrained residual scalar quantization.[2] The two encoded components are then time-multiplexed into the narrowband telephone channel for transmission to the remote location. At the receiver, the channel signals are demultiplexed and the individual components are decoded. The lowpass component is then upsampled and interpolated spatially to restore it to is proper size and the texture component is upsampled temporally to the original frame rate. After the components are restored, they are combined the 2-D nonlinear synthesis section to form the reconstructed video. In the next sections, we take a closer look at the individual operations shown in the block diagram in Figure 1.

3 NONLINEAR SUBBAND DECOMPOSITION

Two particular characteristics of the ultrasound video signal support the idea of using the model-based decomposition. First, if a static tissue is examined, the ultrasound image can be interpreted as the product of a luminance lowpass component, representing the intensity of the ultrasonic wave in the vicinity of the examined tissue, and a constant reflectance component, representing the reflection coefficients associated with the tissue. Second, ultrasound images are typically very noisy. Usually, additive noise models are used to describe the effect of noise in images. Filtering out the noise could enhance the



(a)



(b)

Figure 1: The components of the coding system: (a) sender, (b) receiver

images, but more important it makes the image easier to code. If the noise has a gaussian distribution then a linear filter is optimal for maximizing the signal-to-noise ratio. In our case, however, the goal is to maximize the subjective quality of the lowpass component.

Thus two approaches can be considered: an additive model and a multiplicative model. A model formulation that covers both additive and multiplicative variates is depicted in Figure 2. It is similar in nature to the homomorphic model pioneered by Stockham[3] for the purpose of image enhancement.

The filter $H(\omega)$, shown in Figure 2, is a lowpass filter with a cutoff frequency of $\omega_c = \pi/D_1$. The nonlinear decomposition is then described by the equation

$$x_{n,i,j} = \Psi^{-1}(\Psi(l_{n,i,j}) + \Psi(t_{n,i,j})).$$

This decomposition is equivalently a nonlinear subband decomposition. The nonlinearity $\Psi(\cdot)$ is chosen to be of the form

$$\Psi(x) = \beta x^{\alpha}.$$

For $\beta = 1$ and $\alpha = 1$, $\Psi(\cdot)$ is the identity mapping and we obtain the additive model. For $\alpha = 0.231$, $\Psi(x) \simeq \beta \log(x)$ in the range 0 to 255, and we obtain the multiplicative model.

The parameter β was chosen so that x and $\Psi(x)$ have the same dynamic range, i.e. from 0 to 255. The parameter α was chosen empirically to optimize the subjective performance. Qualitatively, we want the lowpass component to contain as much useful detail as possible, while keeping constant the cutoff frequency of the filter $H(\omega)$. To quantify this criterion, we could try to minimize the difference between $l_{n,i,j}$ and $x_{n,i,j}$ to address the aforementioned goal. Similarly, we could try to minimize the energy in the texture $t_{n,i,j}$. This ensures that the amount of information contained in the texture is not significant. We have measured these quantities for values of α in the range $0.1 \leq \alpha \leq 2$ for a sample set of ultrasound images and the results are summarized in Figure 3. The graph (a) shows the dependency of the mean square difference between $l_{n,i,j}$ and $x_{n,i,j}$ and the graph (b) shows the dependency of the energy of $t_{n,i,j}$ on the parameter α.

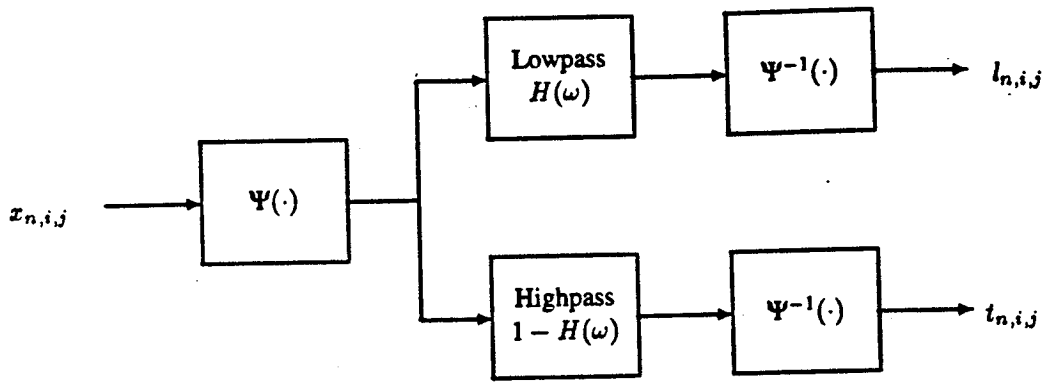
We can see that the two criteria are conflicting, and a compromise between them is needed. The value $\alpha = 0.231$ that implies an approximately logarithmic mapping is in the range of values that provide a good tradeoff between the two criteria. Therefore, the multiplicative model is a reasonable model to use for the encoding of ultrasound images.

The additive and multiplicative models are compared in Figure 6. A sample original ultrasound image is presented together with the reconstructed images obtained by using the additive ($\alpha = 1$) and multiplicative ($\alpha = 0.231$) models. We can see that the multiplicative model has improved subjective appearance.

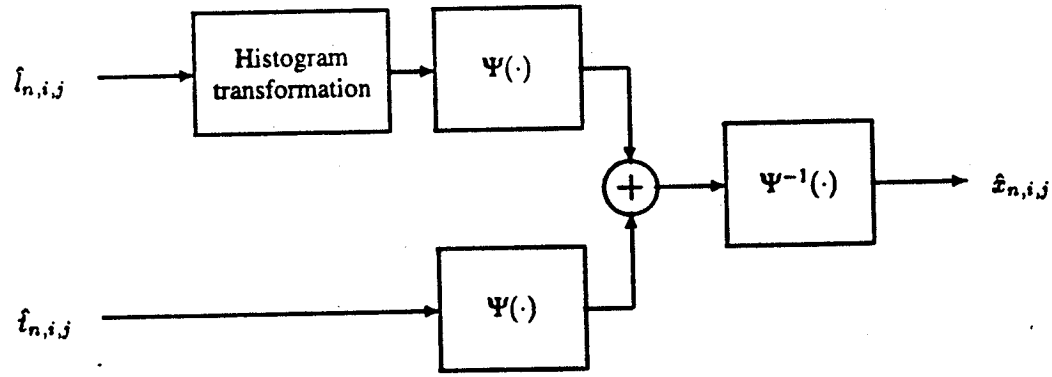
Because this decomposition is similar to the homomorphic luminance-reflectance decomposition introduced by Stockham[3] in the context of image enhancement, we can also hope to be able to introduce some image enhancement capability to the ultrasound images. In fact, the system is constructed with this feature. Unlike Stockham's approach where different gains are imposed on the two components, we perform histogram modification of the lowpass component. This provides greater flexibility for enhancement. The histogram transformation used in this paper is nonlinear and has the profile shown in Figure 4. It was observed experimentally that the features most difficult to preserve during encoding are represented in the low and medium amplitude range of the lowpass component. Thus, contrast modification in this region is expected to enhance perceived quality. Preliminary results indicate that this is true. At this point enhancement results have not been evaluated by medical specialists, but hopefully will be by the time of the conference presentation.

4 LOWPASS COMPONENT CODING

Taking into account the way the lowpass component $l_{n,i,j}$ was obtained, it can be represented by $\Psi(l_{n,i,j})$, which is a bandlimited signal with a cutoff frequency of π/D_1 in both horizontal and vertical directions. Therefore, $\Psi(l_{n,i,j})$ can be



(a)



(b)

Figure 2: The (a) analysis and (b) synthesis subsystems

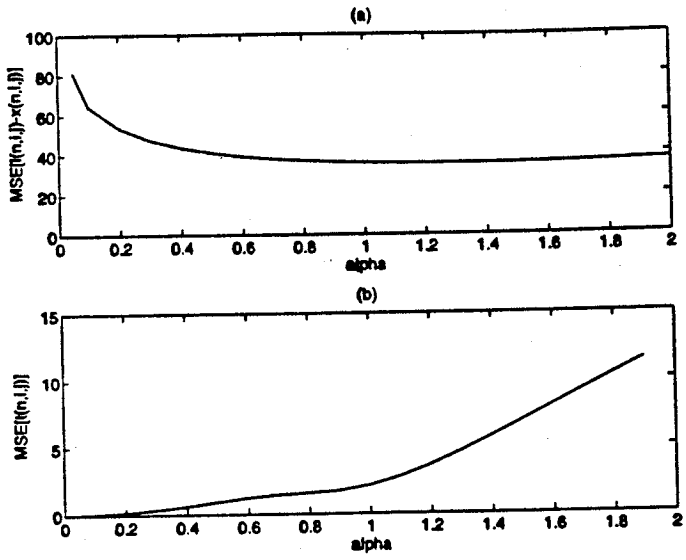


Figure 3: Dependency on α of the two optimality criteria. (a) MSE of $(l_{n,i,j} - t_{n,i,j})$ as a function of α. (b) MSE of $t_{n,i,j}$ as a function of α.

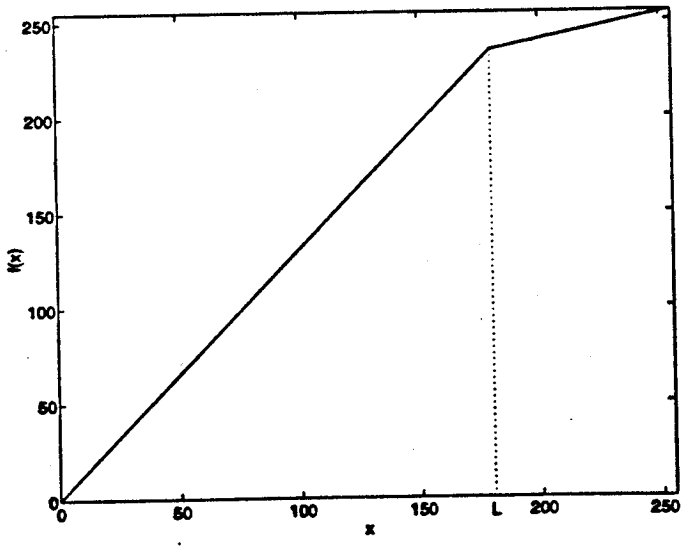


Figure 4: Example of nonlinear histogram transformation for image enhancement.

decimated by D_1 in both horizontal and vertical directions without loss of information due to aliasing. Since the mapping $\Psi(\cdot)$ is one-to-one and onto, $l_{n,i,j}$ can be decimated and reconstructed. By coding the decimated version of $l_{n,i,j}$, the net bit rate can be reduced dramatically.

One of the front-running image coding techniques is subband coding.[4,5] It refers to a broad class of systems where the input is decomposed into subband images and the subband images are coded for transmission or storage. In this work, an new optimized subband image coder is employed.[1] This particular subband coding system consists of decomposing the lowpass component into 16 uniform subbands using the A11 two-band analysis filters introduced in reference [6]. The implementation can be made very efficient computationally by using specially designed recursive filters that require no multiplication operations.[1] The subbands are then quantized using entropy-constrained multistage quantizers with intra-band and inter-band conditioning.

This subband coding system is described in detail in references [1] and [2]. Hence our discussion of this part of the coder is brief. Let it suffice to say that the subband coder is based on encoding each subband pixel (one quantization stage at a time) using conditional entropy coding. The conditioning is based on the quantized symbol values in the local neighborhood of the pixel and in corresponding locations across the subbands. Conditional entropy coding of this form allows statistical dependencies within and across subbands to be used to our advantage.

In addition however, we also extend the conditioning to include corresponding pixels in previous frames. Implementation complexity limits the number of conditioning symbols that can be used practically, which is unfortunate. Therefore only the most statistically important conditioning symbols are used (the precise number being fixed a priori by implementation constraints). For a fixed number of conditioning symbols, an algorithm that finds the location of conditioning symbols such that the overall entropy is minimized is described in.[7]

Conditioning on previous frames is reasonable since there is a lot of correlation between consecutive frames, especially after the noise has been filtered out in the nonlinear subband decomposition stage. This conditioning scheme is described in Figure 5, where only spatial conditioning is depicted. Inter-subband and inter-stage conditioning are not shown in the figure for clarity resons. Solid lines represent intra-frame conditioning and dashed lines represent inter-frame conditioning. Note that this conditioning scheme requires a large number of previous frames to be buffered. However, this is not a big problem in our case, because the frames are small (64 by 64 pixels).

This type of conditioning for cardiology ultrasound video can be used to exploit the fact that the image sequence is quasiperiodic, with a period given by the heartbeat rate. Therefore, we can use conditioning based on symbols from the frame located one heartbeat period before the current frame. A couple of techniques can be used for estimating the heartbeat period. Ideally, we would choose the value that minimizes the average codeword in the current frame. This method is computationally intensive. A simpler method is to use for conditioning the frame that minimizes the difference between itself and the current frame. However, ultrasound machine outputs often provide the EKG signal explicitly. Thus the simplest way is to extract the period directly from the accompanying EKG.

5 TEXTURE CODING

The texture component is a valuable part of the coded signal in the sense that it contributes to the natural appearance of the reconstructed image. However, much of this texture component is just random noise. One can postulate that the texture of the tissue in the examined region is the same for a relatively long period of time, except when the sensor device is in motion, and additive noise contributes most to the rapid time variations in the signal statistics. A simple approach to encode the texture component is to decimate it in the temporal dimension. A texture frame is then only encoded and transmitted once every D_t frames. At the receiver, the same decoded texture frame is used for the synthesis of D_t consecutive frames.

For large values of D_t, this method may produce an unpleasant effect of static texture. In order to reduce this effect and to have a more subjectively realistic decoded video sequence, two consecutively transmitted texture frames may be used

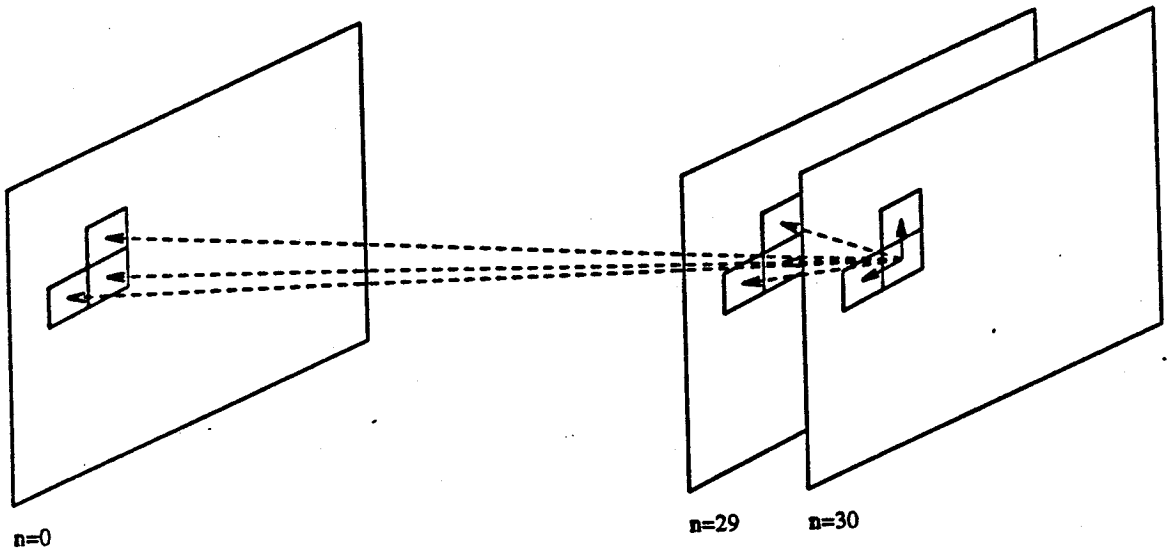


Figure 5: The inter-frame and intra-frame conditioning scheme

alternatively. Alternating between these two texture frames every 1/30 seconds is an improvement but the flicker effects are too strong. Further subjective improvement is achieved by switching between them every two or three 1/30 second periods. Experiments have shown that the best subjective quality is obtained when we switch texture frames every three lowpass component frames.

The same entropy-constrained quantization technique is used to encode the texture. The only difference is that conditioning is realized only with respect to neighboring pixels and neighboring subbands. This is because there is no significant correlation between textures D_2 frames apart.

6 EXPERIMENTS AND RESULTS

If the lowpass component is encoded at R_1 bits per pixel and the texture at R_2 bits per pixel, the overall bit rate in bits per second is given by

$$R = 30\left(\frac{256 \times 256}{D_1^2}R_1 + \frac{256 \times 256}{D_2}R_2\right).$$

The lowpass component spatial decimation factor D_1 can be equal to 8 if the coding system is used for positioning only, 4 if we need diagnostic quality, and 2 or even 1 if we implement a multiresolution system allowing zooming in the area of interest. The texture component temporal decimation factor is in the range 25 to 40.

In Figure 7 we present an original ultrasound image (a), the corresponding lowpass component (b), the reconstructed image (c), and the reconstructed image with contrast enhancement (d). Enhancement has been performed using the histogram transformation depicted in Figure 4 with the parameters $L = 170$ and $f(L) = 210$.

The parameters used for encoding are $D_1 = 4$, $D_2 = 60$, $R_1 = 0.45$, $R_2 = 0.25$. The overall bit rate is then $R = 55kbps + 8kbps = 63kbps$, so this example can be used for transmission over an ISDN line. We have used a uniform 64-subband decomposition, and a quantizer having six stages and two code vectors per stage.

Encoded video segments will be presented at the conference.

7 ACKNOWLEDGMENTS

The authors would like to thank Dr. Casimir Eubig and Dr. William A. Lutin at the Medical College of Georgia for their general input, providing us data, and their evaluation of the coded ultrasound images.

8 REFERENCES

[1] A. Docef, F. Kossentini, W. Chung, and M. Smith, "Multiplication-free subband coding of color images," in *Data Compression Conference*, (Snowbird, UT, USA), Mar. 1995. Submitted for publication.

[2] F. Kossentini, M. Smith, and C. Barnes, "Entropy-constrained residual vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. V, (Minneapolis, MN, USA), pp. 598–601, Apr. 1993.

[3] T. G. Stockham, Jr., "Image processing in the context of a visual model," *Proc. IEEE*, vol. 60, pp. 828–842, July 1972.

[4] J. W. Woods, ed., *Subband Image Coding*. Norwell, MA: Kluwer Academic Publishers, 1991.

[5] P. H. Westerink, *Subband Coding of Images*. PhD thesis, T. U. Delft, 1989.

[6] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1446–1456, Aug. 1991.

[7] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on Image Processing*, July 1994.

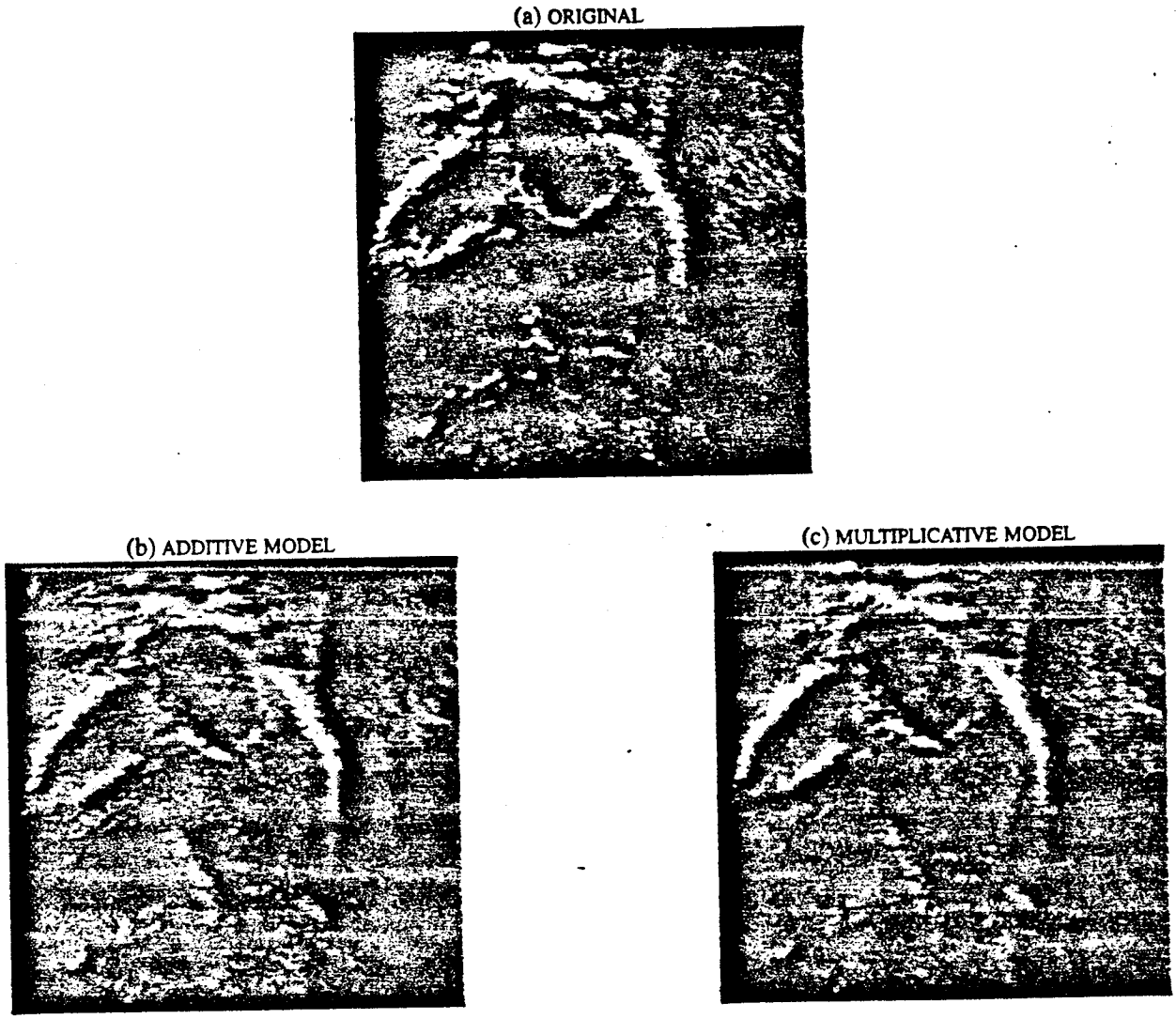
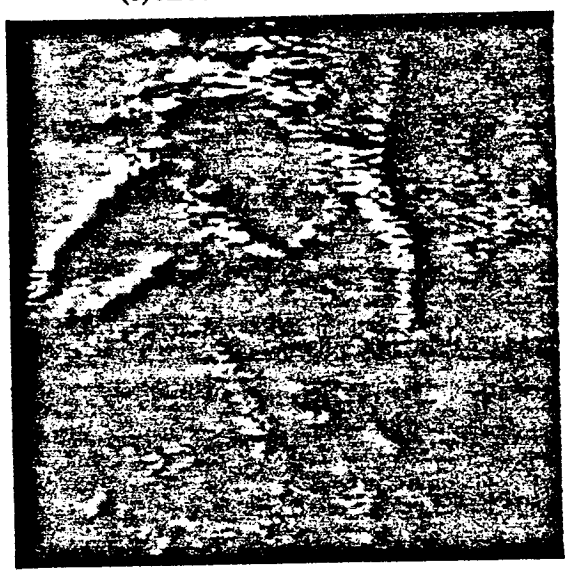


Figure 6: (a) original ultrasound image (b) reconstructed image with additive model (c) reconstructed image with multiplicative model.

(a) ORIGINAL

(b) ORIGINAL LOWPASS COMPONENT

(c) RECONSTRUCTED IMAGE



(d) RECONSTRUCTED IMAGE WITH ENHANCEMENT

Figure 7: Encoded images at $R_1 = 0.45$bpp and $R_2 = 0.25$bpp.

A New Approach to Scalable Video Coding[1]

Wilson C. Chung, Faouzi Kossentini, and Mark J. T. Smith
Digital Signal Processing Laboratory
Georgia Institute of Technology
School of Electrical & Computer Engineering
Atlanta, Georgia 30332-0250, USA

Abstract

This paper introduces a new framework for video coding that facilitates operation over a wide range of transmission rates. The new method is a subband coding approach that employs motion compensation, and uses prediction-frame and intra-frame coding within the framework. It is unique in that it allows lossy coding of the motion vectors through its use of multistage residual vector quantization. Moreover, it provides a rate-distortion-based mechanism for alternating between intra-frame and inter-frame coders. The framework provides an easy way to control the system complexity and performance, and inherently supports multiresolution transmission. Experiments using the standard sequence MISS AMERICA show that the coder significantly outperforms the $p \times 64$ coder (H.261 standard), while still maintaining reasonable complexity.

1 Introduction

Over the past several years many effective approaches to the problem of video coding have been demonstrated. Motion compensation (MC) is the cornerstone of most video coding systems presently in vogue and is the basic mechanism by which temporal redundancies are captured in the current H.261 and MPEG standards. The coding gains achievable by using motion compensated prediction and residual frame coding are well known. Similarly exploiting spatial redundancy within the video frames plays an important part in video coding. DCT-based schemes as well as subband coding methods have been shown to work well for this application. Such methods are typically applied to the MC prediction residual and to the individual frames. In the strategy adopted by MPEG the sequence of video frames is first grouped into blocks of N, where the first frame in the block (the "so-called" I-frame) is coded using an intra-frame DCT coder, while the other $N-1$ frames (called P-frames) are described by motion vectors and coded residuals. Clearly such an approach has been very effective. However, there are notable limitations. In particular, the high level of performance is not maintained over a full range of rates. Most notably, MPEG is not suited for low bit rate applications below 64 kbps. In addition, it is limited in the way it exploits the spatio-temporal variations in a video sequence.

In this paper we introduce a more flexible framework based on a subband representation, motion compensation, and a new class of predictive quantizers. As with MPEG, the concept of I-frame and P-frame is used. However the coding strategy optimizes spatio-temporal coding within subbands. Thus a unique feature of the proposed

[1]This work was supported by a grant from NASA

approach is that it allows for the flexible allocation of bits between the inter-frame and intra-frame coding components of the coder. This is made possible by the use of multistage residual vector quantizers (RVQ) in the subbands and a high order conditional entropy coding scheme that exploits statistical dependencies between motion vectors in the same subband as well as between those in different subbands of the current and previous frames, simultaneously. The new approach, which we develop next, leads to a fully scalable system, with high quality and reasonable complexity.

2 The Video Coder

A high-level description of the framework is shown in the block diagram in Figure 1. Each frame of the input is first decomposed into subbands using a tree-structured recursive analysis filter bank. The filter bank is based on a separable two-band decompositions, which employs allpass polyphase filters as described in [1]. Each of the subbands is then decomposed into blocks of size $W \times H$, where each block $\mathbf{X}$ is encoded using two coders. One is an intra-frame coder, denoted I-subband coder; the other is an MC-predictive coder, denoted P-subband coder. The encoder that produces the minimum Lagrangian distortion is selected, and side information is sent to the decoder indicating which of the two coders was chosen.

As will be explained later, both the I-subband and P-subband coders employ high order entropy coding which exploits dependencies between blocks within and across subbands. Since only one coder is chosen at a time, symbols representing some of the coded blocks in either coder will not be available to the decoder. Thus, both the encoder and decoder must estimate these symbols. More specifically, suppose that the I-subband coder is chosen. Then, the coded block is predicted and quantized using the P-subband coder at both the encoder and decoder. If the P-subband coder is chosen, then the motion-compensated reconstructed block is quantized using the I-subband coder, also at both the encoder and decoder. In this case, the additional complexity is relatively small because the the motion-compensated reconstructed blocks are available at both ends, and only I-subband quantization is performed. In practice, most of the blocks are coded using the P-subband coder. Thus, the overall additional complexity is relatively small.

Figure 2 shows a typical generic coder (GC) structure for the I-subband coder. An input block is first divided into sub-blocks or vectors, which are then quantized using a multistage residual vector quantizer (GQ) [2] and entropy coded using a binary arithmetic coder (BAC) [3, 4] that is specified by a finite-state machine (FSM) high order statistical model [5, 6].

Figure 3 shows the structure of the P-subband coder. A full-search block matching algorithm (BMA) using the mean absolute distance (MAD) is used to estimate the motion vectors. Since the BMA does not necessarily produce the true motion vectors, we employ a thresholding technique for improving the rate-distortion performance. Let d_{min} be the minimum MAD associated with a block to be coded. Also, let $T > 1.0$ be a threshold, which is empirically determined from the statistics of the subband being coded. We then choose all motion vectors with associated MADs d_i

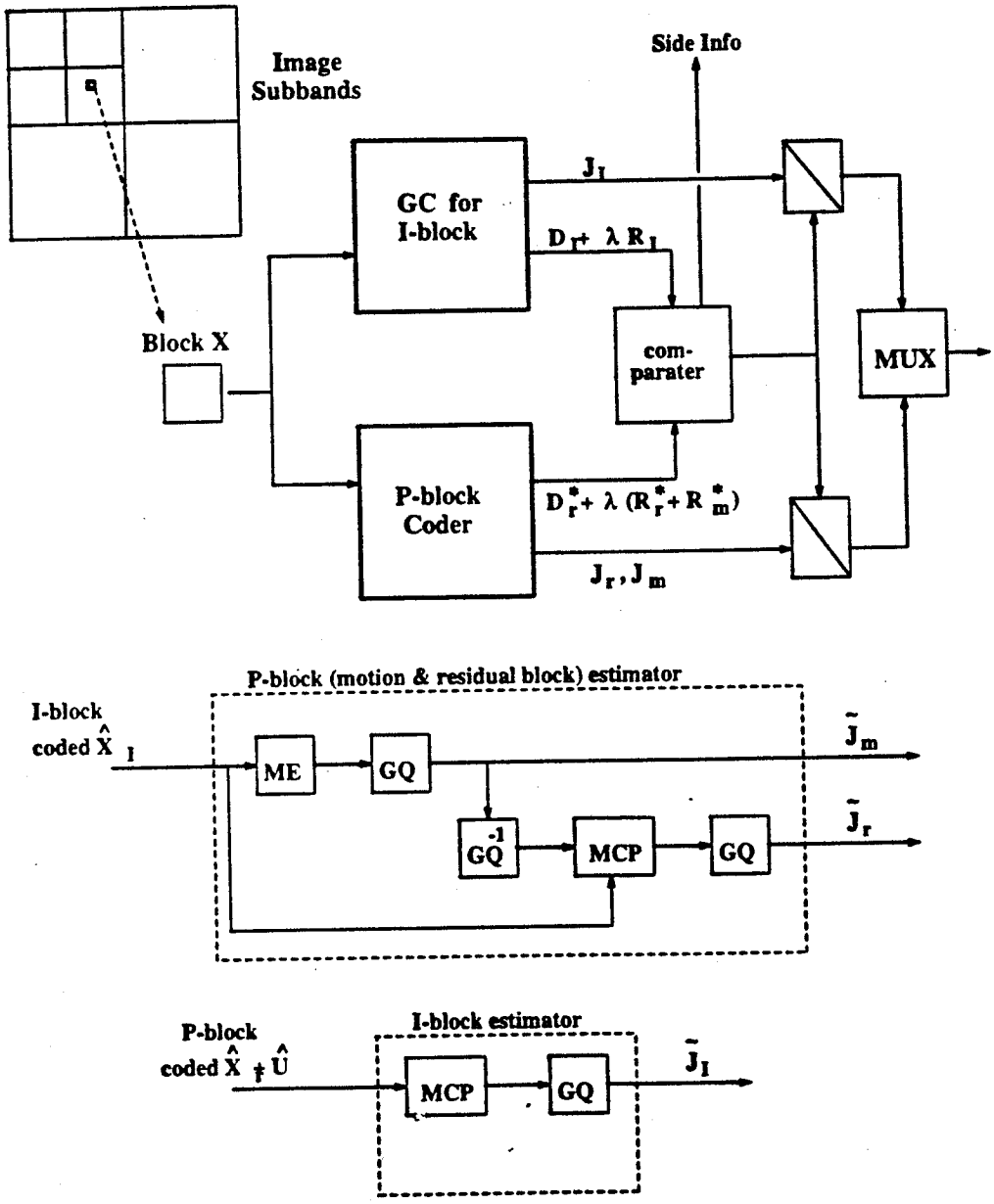


Figure 1: Block-level diagram for the proposed video coder.

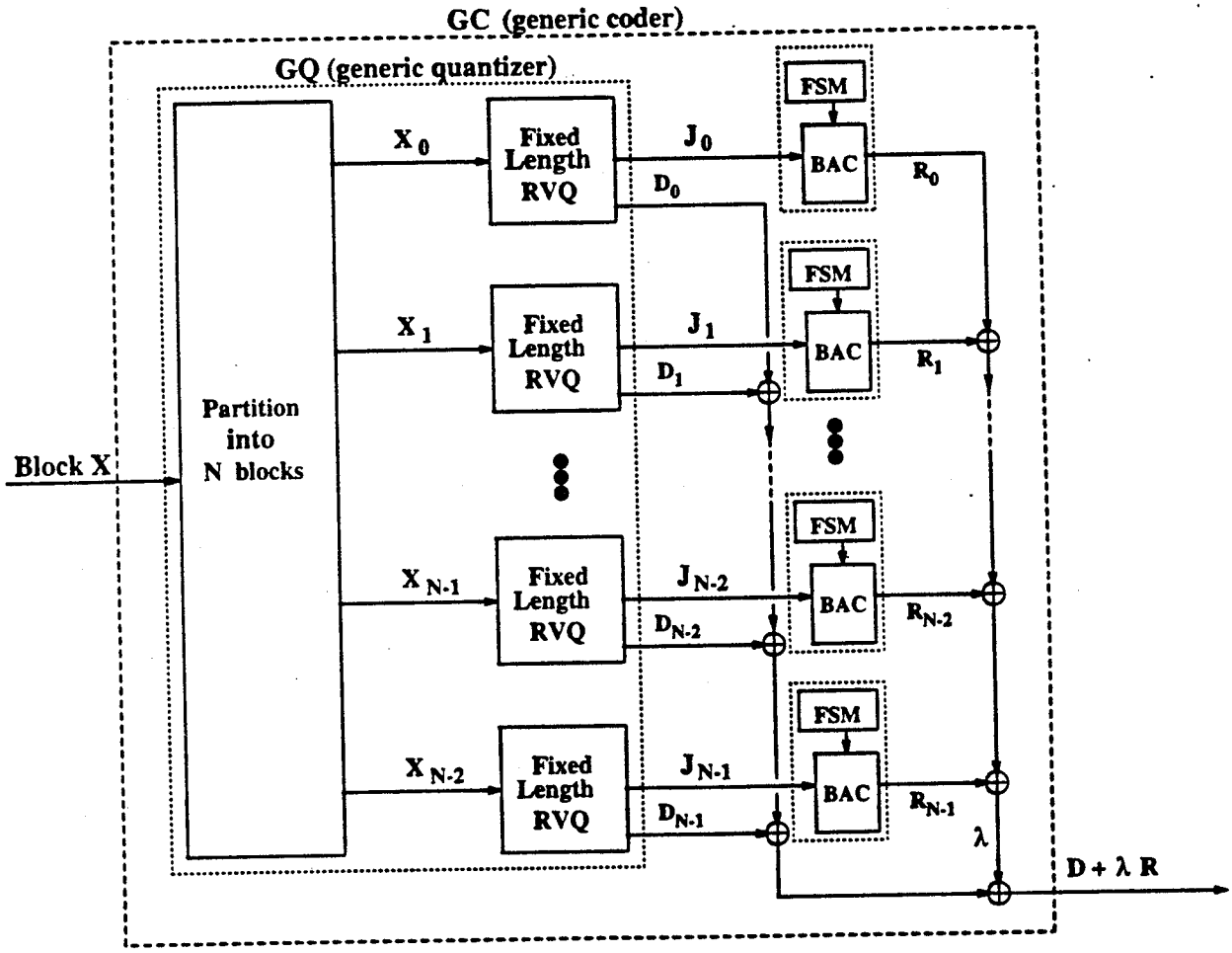


Figure 2: A typical generic coder (GC).

that satisfy $\frac{d_i}{d_{min}} < T$. When the motion estimate is accurate, the number M of candidate motion vectors is more likely to have a value of 1. In cases where the video signal undergoes sudden changes (e.g., zoom, occlusion, etc ...) that cannot be accurately estimated using the BMA algorithm, a large number of motion vectors can be chosen as candidates, thereby leading to a large complexity. Thus, we limit the number M of candidate motion vectors to a relatively small number M_{max} (e.g., 3 or 4).

In many conventional video subband coders as well as in MPEG, differential entropy coding of motion vectors is employed. Since motion vectors are usually slowly varying, the motion bit rate can be further reduced by exploiting dependencies not only between previous motion vectors within and across the subbands but also between the vector coordinates. For this purpose, we employ a high rate multistage residual 2-dimensional vector quantizer cascaded by a BAC specified by the FSM statistical model described later. After the M candidate motion vectors are vector quantized, M motion-compensated prediction blocks are generated, and corresponding residual blocks are computed. The encoding of the residual block is done in the same manner as for the original block.

An important part of the encoding procedure is the decision where either the I-subband or the P-subband coders must be chosen for a particular block. Figure 1 shows the encoder procedure. Let R_I and D_I be the rate and distortion associated with I-subband coding the block $\mathbf{X}$, respectively. Also, let R_M^m, $1 \leq m \leq M$ be the rate required by the 2-dimensional motion coder for the mth candidate motion vector and (R_r^m, D_r^m) be the rate-distortion pair for the corresponding residual coder. Assuming λ is the Lagrangian parameter that controls the rate-distortion tradeoffs, the I-subband coder is chosen if

$$D_I + \lambda R_I < D_r^m + \lambda(R_r^m + R_M^m), form = 1, \ldots, M.$$

Otherwise, the P-subband coder that leads to the lowest Lagrangian is chosen.

Statistical modeling for entropy coding the output of the I-subband, motion vector, and residual coders consists of first selecting, for each subband, N conditioning symbols from a region of support representing a large number of neighboring blocks in the same band as well as other previously coded bands. Then let F be a mapping that is given by

$$u = F(s_1, s_2, \ldots, s_N),$$

where $s_1, s_2, \ldots, s_N$ are the N selected conditioning symbols and $u \in \mathcal{U}$ represents the state of the stage entropy coder. The mapping F converts combinations of realizations of the conditioning symbols to a particular state. For each stage in each subband, a mapping F^* is found such that the conditional entropy $H(J|U)$, where J is the stage symbol random variable and U is the state random variable, are minimized subject to a limit on complexity, expressed in total number of probabilities that must be computed and stored. The tree-based algorithms described in [5, 6] are used to find the best value of N subject to a limit T_1 on the total number of probabilities. The PNN algorithm [7], in conjunction with the generalized BFOS algorithm [8], is then used to construct tables that represent the best mappings F^* for each stage

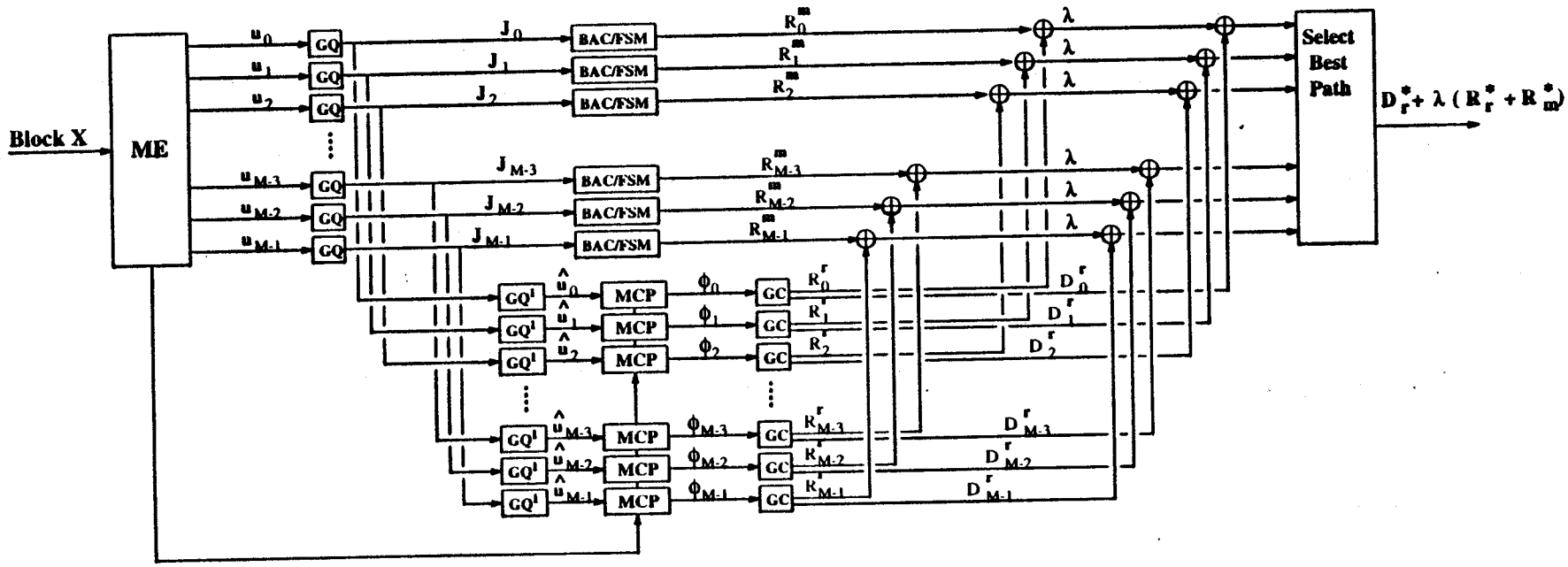


Figure 3: A P-subband (MC-Predictive) coder.

entropy coder subject to another limit T_2 ($T_2 << T_1$) on the number of probabilities. Note that the number T_1 controls the tradeoffs between entropy and complexity of the PNN algorithm. The output of the RVQs are eventually coded using a adaptive binary arithmetic coders (BACs) [3, 4] as determined by the FSM model. BAC coders are very easy to adapt and require small complexity. They are also naturally suited for this framework because experiments have shown that a stage codebook size of 2 usually leads to a good balance between complexity and performance and provides the highest resolution in a progressive transmission environment.

3 PRACTICAL DESIGN ISSUES

In both the I-subband and P-subband coders, the multistage RVQs and associated stage FSM statistical models are designed jointly using an entropy and complexity-constrained algorithm, which is described in [5, 6]. The design algorithm iteratively minimizes the expected distortion $E\{d(\boldsymbol{X}, \hat{\mathbf{X}})\}$ subject to a constraint on the overall entropy of the stage FSM models. The algorithm is based on a Lagrangian minimization and employs a Lagrangian parameter λ to control the rate-distortion tradeoffs. The overall FSM statistical model, or the sequence of stage FSM models, enables the design algorithm to jointly optimize the entropy encoders and decoders subject to a limit on complexity. To substantially reduce the complexity of the design algorithm, only independent subband fixed-length encoders and decoders are used. However, the RVQ stage quantizers in each subband are jointly optimized through dynamic M-search [9], and the decoders are jointly optimized using the Gauss-Seidel algorithm [2].

To achieve the lowest bit rate, the FSM models used to entropy code the output of the RVQs should be generated on-line. However, this requires a two-pass process where statistics are generated in the first pass, and the modeling algorithm described above is used to generate the conditional probabilities. These probabilities must then be sent to the BAC decoders so that they can track the corresponding encoders. In most cases, this requires a large complexity. Moreover, even by restricting the number of states to be relatively small (such as 8), the side information can be excessive. Therefore, we choose to initialize the encoder with a generic statistical model, which we generate using a training video sequence, and then employ dynamic adaptation [3] to track the local statistics of the motion flow.

The proposed coder has many practical advantages, due to both the subband structure and the multistage structure of RVQ. For example, multiresolution transmission can be easily implemented in such a framework. Another example is error correction, where the more probable of the two stage code vectors is selected if an uncorrectable error is detected. Since each stage code vector represents only a small part of the coded vector, this will not significantly affect the reconstruction or the FSM statistical models.

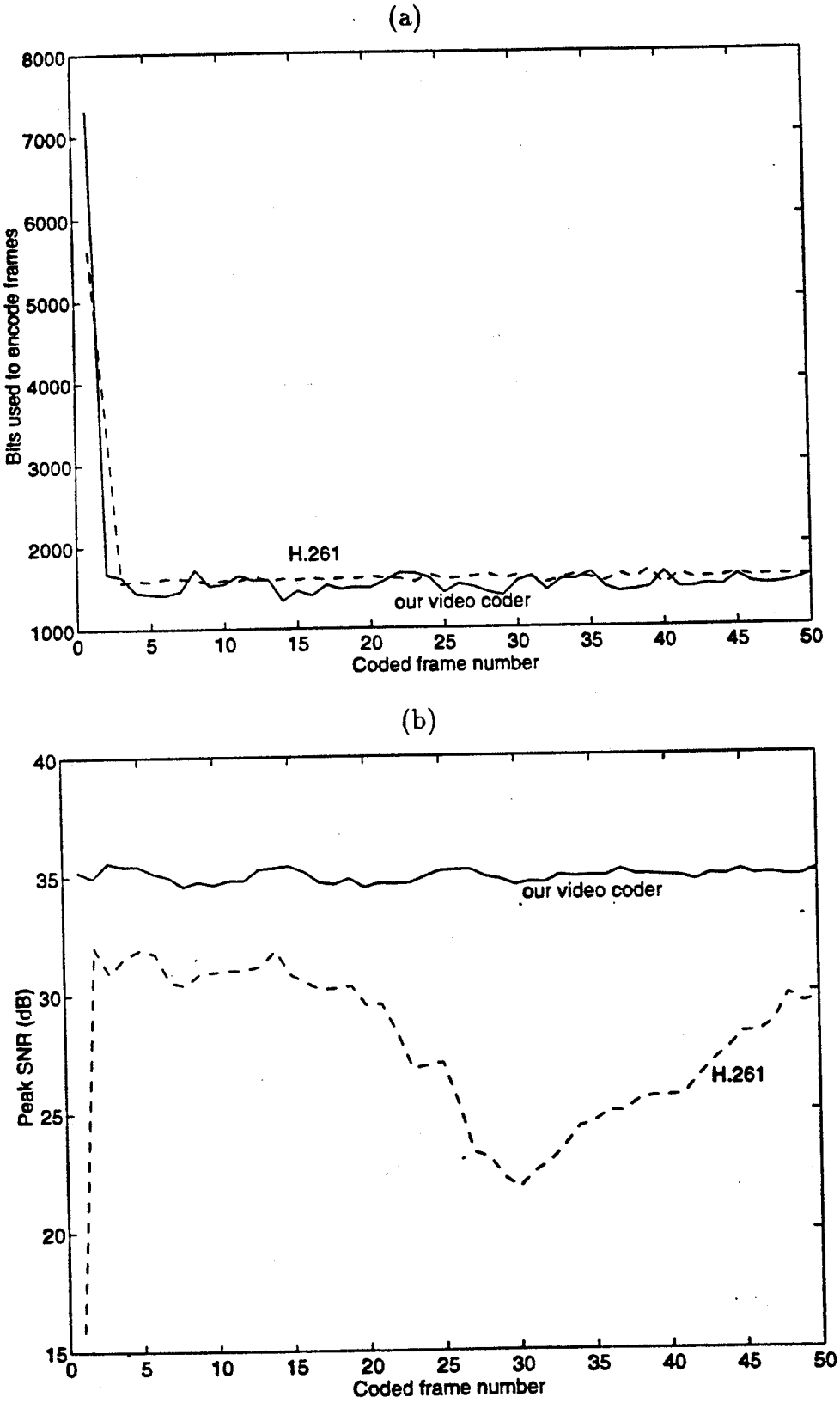


Figure 4: (a) Overall rate usage and (b) Peak SNR performance for the proposed coder.

4 EXPERIMENTAL RESULTS

To demonstrate the performance of the new entropy coding technique, we designed a low bit rate coder for QCIF (176 x 144 x 10 Hz) YUV color video sequences. We use the full-search algorithm for estimating the block motion vectors with mean absolute difference (MAD) criterion for the measure of the match between two blocks. In practice, it is found that the MAD error criterion works satisfactorily well [10]. Motion estimation is performed only on the Y luminance component and the estimated motion vector field is subsequently decimated spatially in horizontal and vertical directions, and in magnitude for the use of motion compensation of U and V chrominance signals. Figure 4 (a) shows the overall bit rate usage and Figure 4 (b) shows the Peak SNR result of our coder in comparison with H.261 standard for 50 frames of the color test video sequence MISS AMERICA. The average bit rate is approximately 15.98 kbps and the average Peak SNR is 35.13 dB.

References

[1] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1446–1456, Aug. 1991.

[2] F. Kossentini, M. Smith, and C. Barnes, "Necessary conditions for the optimality of variable rate residual vector quantizers," *Submitted to Transactions on Information Theory in June 1993. Revised in May 1994.*

[3] G. G. Langdon and J. Rissanen, "Compression of black-white images with arithmetic coding," *IEEE Transactions on Communications*, vol. 29, no. 6, pp. 858–867, 1981.

[4] W. B. Pennebaker, J. L. Mitchel, G. G. Langdon, and R. B. Arps, "An overview of the basic principles of the Q-coder adaptive binary arithmetic coder," *IBM J. Res. Dev.*, vol. 32, pp. 717–726, Nov. 1988.

[5] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on Image Processing*, July 1994.

[6] F. Kossentini, W. Chung, and M. Smith, "Subband image coding with jointly optimized quantizers," in *Abstract Submitted to IEEE Int. Conf. Acoust., Speech, and Signal Processing*, (Detroit, MI, USA), Apr. 1995.

[7] W. H. Equitz, "A new vector quantization clustering algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, pp. 1568–1575, October 1989.

[8] E. A. Riskin, "Optimal bit allocation via the generated BFOS algorithm," *IEEE Trans. on Information Theory*, vol. 37, pp. 400–402, Mar. 1991.

[9] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114–116, July 1994.

[10] M. I. Sezan and R. L. Lagendik, *Motion Analysis and Image Sequence Processing.* Boston: Kluwer Academic Publishers, 1992.

[9] F. Kossentini and M. Smith, "A fast searching technique for residual vector quantizers," *Signal Processing Letters*, vol. 1, pp. 114–116, July 1994.

[10] M. I. Sezan and R. L. Lagendik, *Motion Analysis and Image Sequence Processing.* Boston: Kluwer Academic Publishers, 1992.

RATE-DISTORTION-CONSTRAINED STATISTICAL MOTION ESTIMATION FOR VIDEO CODING

Wilson C. Chung, Faouzi Kossentini, and Mark J. T. Smith

Digital Signal Processing Laboratory
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250, USA

ABSTRACT

A rate-distortion-constrained statistical motion estimation algorithm is presented here that leads to improvements in subband video coding. The main advantages of the algorithm is that it requires a relatively small number of computations, produces a much smoother motion field, and employs a more effective measure of performance that the conventional mean absolute difference or mean squared error. The proposed algorithm circumvents problems in the motion compensation loop such as illumination variations, noise, and occlusions, by providing a mechanism for alternating between intra-frame and residual coding. Experimental results demonstrate that the corresponding video coder outperforms the H.263 in terms of motion vector search complexity and overall bit rate at the same reproduction quality.

1. INTRODUCTION

In conventional video coding systems, block matching algorithms (BMAs) are often used for motion estimation to remove temporal redundancies [1, 2, 3]. Such algorithms form the foundation for many video coders and are part of the H.261, H.263, and MPEG standards [4, 5, 6, 7], mainly because they are relatively simple in concept and design, but also because they tend to work reasonably well.

A disadvantage of BMAs, in general, is that their performance is sensitive to illumination changes, noise, occlusion, and reconstruction quality of previously coded frames. Motion vector estimates often do not correspond to physical motion in the video scene. Even where motion does not exist, BMAs produce an estimate. This can lead to a rough motion field, where many motion vectors carry little useful information, yet are very difficult to encode. Moreover, since a mean squared error or mean absolute difference distortion measure is usually used as the matching criterion, the motion vector estimate does not necessarily lead to the best rate-distortion performance [8, 9, 10]. To address some of these problems, the MPEG-2 standard, for example, provides a mechanism for alternating between intra-frame and inter-frame coders.

In this paper, we introduce a rate-distortion constrained statistical motion estimation algorithm that not only requires a level of complexity that is comparable to that of the fastest BMAs but also solves most of the above problems, thereby leading to a more consistent motion field. The algorithm is used for motion estimation on the subband level [11, 12], where high order entropy-constrained residual scalar quantization is employed for coding both the original and residual subbands. The proposed algorithm exploits the natural motion field smoothness that tends to exist spatially, temporally, and across subbands. It selects motion vectors based on the current behavior of the motion field and also based on the performance of the residual coder, which is also the ultimate objective performance measure of the video coder. Although the proposed algorithm is presented in the context of subband video coding, its underlying principles can also applied in other contexts.

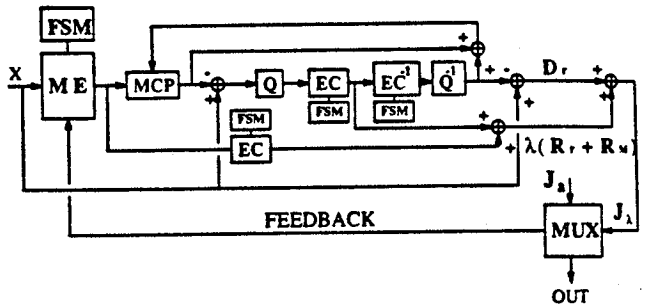


Figure 1: Block diagram of rate-distortion-constrained statistical motion estimation.

2. THE PROPOSED MOTION ESTIMATION ALGORITHM

The proposed motion estimation algorithm is illustrated in Fig. 1. Each frame of the video sequence is decomposed into M subbands using a uniform subband decomposition structure. The algorithm is applied to each subband independently, using information from previously coded subbands (see Fig. 2). First, sufficiently large block and search region sizes are chosen for each subband. All of the possible motion vectors in the search region are then divided into clusters or rectangular regions. This is illustrated in Fig. 3, where each black dot denotes a motion vector location. The search region shown in the figure corresponds to ± 4 displacements for each of the two coordinates. During the

This work was supported in part by the Joint Services Electronics Program (JSEP).

0-8186-7310-9/95 $4.00 © 1995 IEEE

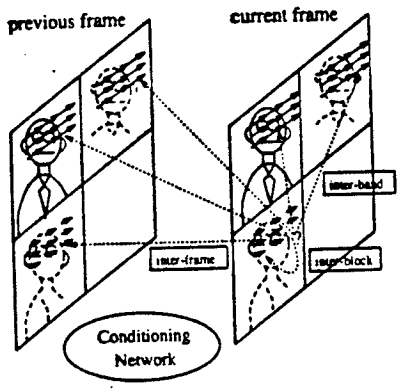


Figure 2: Inter-frame, inter-subband, and intra-subband dependencies for motion vectors.

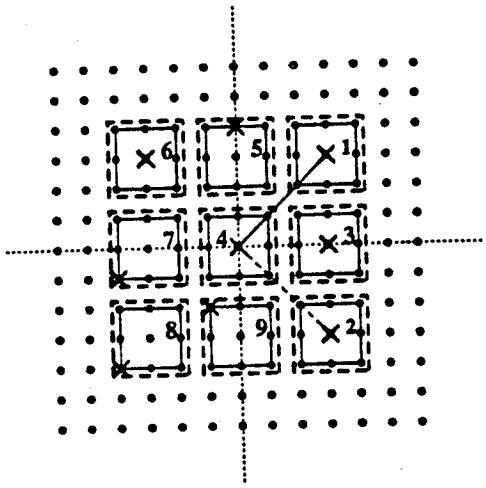


Figure 3: An example of first layer and second layer passes in intra-band motion estimation.

design process, conditional probabilities are generated for each of the rectangular regions, and these probabilities are grouped into tables, each corresponding to a conditioning state. The states are derived based on previously coded motion vector in a subband-spatial region of support. In other words, a high order statistical model that is driven by a finite-state machine (FSM) is built that exploits statistical dependencies in the motion field between motion vectors within the current subband as well as between subbands in both the same frame and previous frames. Complexity reduction techniques described in [13] allow us to use a sufficiently large conditioning subband spatio-temporal region of support, yet produce only a small number of conditioning states. Since the motion field is usually quasi-stationary, adaptation is used during the encoding procedure, but the conditioning network is kept fixed for a larger number of frames.

Given a conditioning state, the algorithm, illustrated in Fig. 3, performs two passes. In the first layer pass, the motion vector (solid line terminated by $\times$) with largest probability p_i contained in the rectangular region with the largest conditional probability p_j (i.e. region 1 in Fig. 3) is selected first as the candidate motion vector. A high order entropy-constrained residual coder [13] is then applied to the difference between the original block and the motion-compensated prediction block, producing a rate R_r and a distortion D_r, as shown in Fig. 1. Next, we compute the Lagrangian $J_\lambda = D_r + \lambda(R_m + R_r)$, where R_m is the motion vector bit rate, set here for simplicity to the sum of conditional self-information components [1] $R_m = -\log_2(p_i) - \log_2(p_j)$. Let J_a be the current running average Lagrangian and T_1 be a threshold[2] that determines the tradeoffs between complexity and rate-distortion performance. If $J_\lambda \leq T_1(J_a)$, then the selected motion vector is accepted and encoding is terminated for that block by sending motion and residual encoded bits to the channel. At this point, practically no computations have been performed for the estimation procedure. All multiplies/adds performed would have been needed for encoding subsequently. If $J_\lambda > T_1(J_a)$, then the selected motion vector is rejected and a signal is fed back to the motion estimator, where the most probable motion vector located in the region with the second largest conditional probability p_j is selected as an alternative candidate. This is indicated by the dashed line terminated by $\times$ in region 2 of Fig. 3. This procedure is repeated until either the above condition is met, when encoding is aborted, or all regions are exhausted. In cases where little or no motion exists in the video scene, encoding is aborted in the early stages of the first layer pass. However, in cases where the video signal undergoes sudden changes (e.g., zoom, occlusion, illumination), accurate motion vectors cannot be predicted based on the probabilities in the model because no *a priori* information about sudden motion variations is available. As a result, an inaccurate motion vector predicted by the statistical model will generally lead to an increase in the Lagrangian value J_λ. In such cases, a second layer pass is employed.

In the second layer pass, the lowest Lagrangian J_λ^* is compared to J_a. If $J_\lambda^* > T_2(J_a)$, where T_2 is a threshold, whose best value is found experimentally to be between 2.0 and 3.0, then the algorithm exits. Otherwise, the region that led to the lowest Lagrangian is again considered, where other less probable motion vectors belonging to the same region are chosen as candidates. The algorithm proceeds by applying the same procedure as in the first layer pass. In other words, for the next most probable motion vector, the new motion-compensated prediction block is computed, and the same entropy-constrained residual scalar coder is applied to the corresponding residual block. The same procedure is repeated until the proper condition is satisfied, or all regions are exhausted. Finally, in the case where the algorithm exits the two passes without yielding any "good" motion vector candidate, the lowest Lagrangian produced during both passes is compared to that of the intra-frame coder, and the coder leading to the lower value is used. Details of the rate-distortion-based mechanism, by which a particular coder is chosen, as well as a complete description of the residual coder can be found in [12].

[1] Note that by storing $-\log_2(p)$ instead of a probability p, no $\log_2$ operations need to be performed.

[2] The best value of T_1 is determined experimentally, and is usually between 1.0 and 1.5.

3. ADVANTAGES OF THE PROPOSED ALGORITHM

At first glance, the proposed motion estimation algorithm seems quite complicated. However, experimental results show that, depending on the bit rates of operation and the contents of the video scene, our algorithm stops after the first stage of the first layer pass, which requires practically no computations, more than 90% of the time. Moreover, the algorithm completes both passes in only approximately 2% of the cases. Besides its computational advantage, the proposed algorithm has several other features worthy of mention. First, it efficiently and effectively exploits dependencies between motion vectors by using a two-layer, region-based and vector-based, statistical model. Second, it improves the consistency of the spatio-temporal smoothness of the motion field and reduces sensitivity of the estimation by favoring the most probable candidates. To illustrate this, Figures 4 (a) and (b) show the comparison between the motion fields resulted from our algorithm and the full-search BMA. The increased smoothness observed in Fig. 4 (b) indicates a lower entropy. Not only it is easier to encode the motion vectors in Fig. 4 (b), but the subjective quality of the reconstructed video frames is also better. Another advantage of our algorithm is that it directly embeds the residual coder into the estimation loop, thereby potentially leading to overall better rate-distortion performance. Finally, note that the number of computations required by the proposed motion estimation algorithm is variable, and depends on the content of the video scene. In variable length video coders (such as MPEG), this can be incorporated into the buffering schemes already being used in bit rate control.

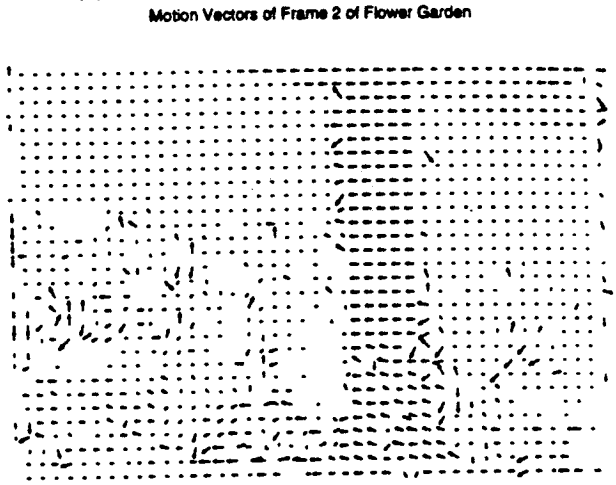
4. EXPERIMENTAL RESULTS

In the experiments, the QCIF version of the MISS AMERICA sequence is used for the test sequence. To compare the performance of our video coder with the current technology for low bit rate video coding (i.e., below 64 kbits/sec), we used the software simulation model of the new H.263 standard obtained from *ftp://bonde.nta.no/pub/tmn* [14].

The subband decomposition is a uniform 2×2 exact reconstruction analysis/synthesis system. We chose the recursive filter banks for their computational efficiency. The block size for searching the motion vector candidates in our algorithm is 4×4. A search region of ± 4 pixels in both spatial directions is chosen. The threshold values T_1 and T_2 are set to be 1.2 and 3.0 respectively. The current running average Lagrangian J_a is computed based on the previous four Lagrangian J_λ values in order to make the algorithm more adaptive. All the motion vectors throughout the experiments are at whole pixel accuracy. Motion estimation is performed only for the luminance component and the estimated motion vector field was subsequently used for the motion compensation of the chrominance signals. The target bit rate is set to be approximately 16 kbits/sec.

Fig. 5 (a) and (b) show the bit rate usage and the PSNR coding performance of our coder and the H.263 standard for 50 frames of the luminance component of the color test sequence MISS AMERICA. We fixed the PSNR and compared the corresponding bit rates required by both coders.

(a) MOTION FIELD FOR FULL-SEARCH BMA



(b) MOTION FIELD FOR OUR ALGORITHM

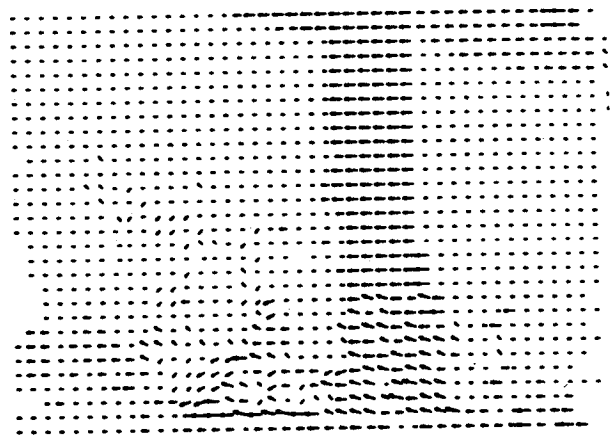


Figure 4: Motion vector field obtained by (a) FS-BMA and (b) our algorithm on the low-pass subband of FLOWER GARDEN (SIF format) Frame No. 2. We used a ± 4 pixel search region with block size of 4×4.

While the average PSNR is approximately 39.4 dB for both coders, the average bit rate for our coder is only 13.245 kbits/sec as opposed to 16.843 kbits/sec for the H.263 standard. To achieve the same PSNR performance, our coder requires only 78% of the overall bit rate of the H.263 video coder.

Finally, to illustrate the computational reduction in motion estimation, we show the comparison in terms of number of matches required for our algorithm and the full-search BMA. For the FS-BMA with the same ± 4 search region in a subband frame, the number of matches required is $22 \times 18 \times 81 = 32076$, and $4 \times 4 = 16$ MAD calculations for each corresponding match. Our algorithm requires at most $22 \times 18 \times (9 + 9 - 1) = 6732$ matches with a table look-up of codebook size 9 with no MAD calculations. For example, in Frame 41 of the MISS AMERICA sequence, 305 out

of $22 \times 18 = 396$ matches are found in the first stage of the first layer pass. Furthermore, no match is found to exhaustively search all stages in both first and second layer passes. The algorithm uses a total of 1202 matches in comparison with 32076 matches that would have been required by the FS-BMA, and our algorithm requires no MAD calculations.

In conclusion, our algorithm outperforms the current H.263 standard by more efficient utilization of bit rates given an image quality. In terms of search complexity in motion estimation, our algorithm is able find a better motion vector field in a rate-distortion sense and requires a fraction of the computation in comparison to the full-search BMA.

5. REFERENCES

[1] H. Musmann, "Advances in picture coding," *Proc. of the IEEE*, vol. 73, pp. 523–548, Apr. 1985.

[2] Q. Wang and R. J. Clarke, "Motion estimation and compensation for image sequence coding," *Signal Processing: Image Communication*, no. 4, pp. 161–174, 1992.

[3] K. I. T. Koga, A. Hirano, Y. Iijima, and T. Ishiguro., "Motion-compensated interframe coding for video conferencing," in *Proc. NTC 81*, (New Orleans), pp. G5.3.1–G5.3.5, Dec. 1981.

[4] "Video codec for autiovisual services at p × 64 kbits/s; Recommendation H.261." The International Telegraph and Telephone Consultant Committee, 1990.

[5] Motion Picture Experts Group, ISO-IEC JTC1/SC29/WG11/602, "Generic Coding of Moving Pictures and Associated Audio," *Recommendation H.262 ISO/IEC 13818-2, Committe Draft*, Seoul, Korea, Nov. 5, 1993.

[6] D. Anastassiou, "Current status of the MPEG-4 standardizaton effort," in *SPIE Proc. Visual Communications and Image Processing*, vol. 2308, pp. 16–24, 1994.

[7] D. Le Gall, "MPEG: a video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, pp. 46–58, Apr. 1991.

[8] B. Girod, "Rate-constrained motion estimation," in *SPIE Proc. Visual Communications and Image Processing*, vol. 2308, pp. 1026–1034, 1994.

[9] D. T. Hoang, P. M. Long, and J. S. Vitter, "Explicit bit minimization for motion-compensated video coding," in *IEEE Data Compression Conference*, (Snowbird, UT, USA), pp. 175–184, Mar. 1994.

[10] G. J. Sullivan., "Multi-hypothesis motion compression for low bit-rate video coding," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. V, pp. 437–440, 1993.

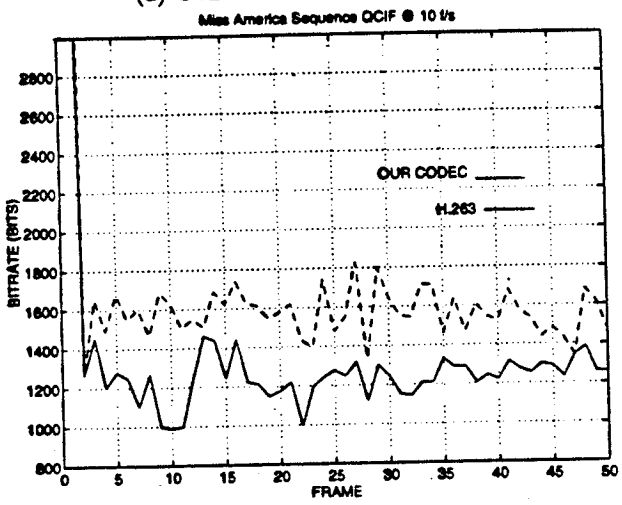
[11] H. Gharavi, "Subband coding algorithms for video applications: Videophone to HDTV-conferencing," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 1, pp. 174–183, June 1991.

[12] W. Chung, F. Kossentini, and M. Smith, "A new approach to scalable video coding," in *IEEE Data Compression Conference*, (Snowbird, UT, USA), Mar. 1995.

[13] F. Kossentini, W. Chung, and M. Smith, "Image coding using high-order conditional entropy-constrained residual VQ," in *International Conference on Image Processing*, (Austin, Texas), Nov. 1994.

[14] Telenor Research, "TMN (H.263) encoder / decoder, version 1.4a, *ftp://bonde.nta.no/pub/tmn*," *TMN (H.263) codec*, may 1995.

(a) OVERALL BIT RATE USAGE



(b) OVERALL CODING RESULTS

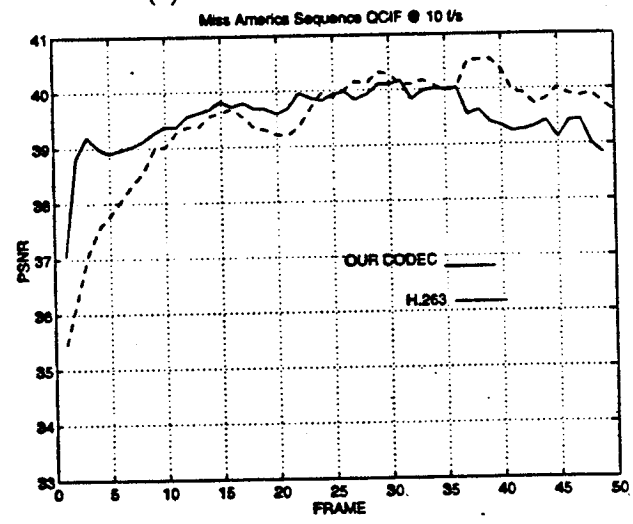


Figure 5: The comparison of overall performance — (a) bit rate usage, and (b) PSNR quality, — of our video coder with H.263 standard for MISS AMERICA QCIF sequence at 10 frames/sec. Not shown in (a) are the values 5465 bits and 7381 bits used by intra-frame (Frame 1) of our coder and H.263 standard respectively. Only the PSNR of Y luminance frames are shown in (b).